RADIO ENGINEERING and ELECTRONIC PHYSICS

English Edition of

РАДИОТЕХНИКА И ЭЛЕКТРОНИКА

Published by the American Institute of Electrical Engineers
with the aid of a grant from the National Science Foundation

No.11 November 1961

Translated and Produced by Royer and Roger, Inc.

AMERICAN INSTITUTE OF ELECTRICAL ENGINEERS

Established 1884

345 East Forty-Seventh Street
New York 17, N.Y.

The English edition of RADIO ENGINEERING AND ELECTRONIC PHYSICS is published by the American Institute of Electrical Engineers with the aid of a grant from the National Science Foundation. Also published under the same arrangement are the Russian electronic journals RADIO ENGINEERING and TELECOMMUNICATIONS.

RADIO ENGINEERING AND ELECTRONIC PHYSICS

(РАДИОТЕХНИКА И ЭЛЕКТРОНИКА)

Publication of the Institute of Radio Engineering and Electronic Physics,
Academy of Sciences of the USSR

Translated and Produced
by
Royer and Roger, Inc.

Translation Editor: Herbert Dern, Columbia University

Subscriptions to Radio Engineering and Electronic Physics should be sent to AIEE
Special Subscription Department
41 East 28th Street, New York 16, New York

1961 Subscription rates:

	$	£
Individuals	28.50	10
Libraries, institutes, govt. agencies	57.00	20

12 issues per annum comprising approximately 1900 pages

TUNNEL DIODES IN AMPLIFIER CIRCUITS (REVIEW)

Yu.I. Vorontsov, K.S. Rzhevkin

We report the main theoretical and experimental results of research on amplification, frequency, noise, and nonlinear properties of tunnel-diode amplifiers. The most typical amplifier designs for various frequency bands are described. Certain ideas are advanced on the prospects of further developments in tunnel-diode amplifiers.

INTRODUCTION

Three years ago Esaki [1] published the first communication on the development of the new semiconductor device, the tunnel diode.

During the time elapsed since, many papers were published on the tunnel effect in p-n junctions [2-6] and on questions connected with the application of tunnel diodes in various radio circuits [7-10, 50, 51]

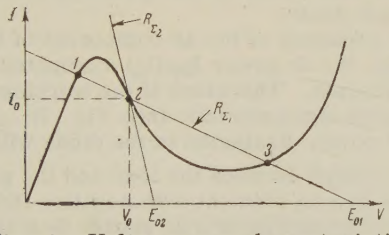


Fig. 1. Volt-ampere characteristic of tunnel diode

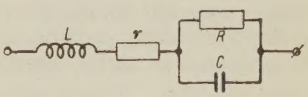


Fig. 2. Equivalent circuit of tunnel diode for alternating signals

As is well known, the tunnel diode is a narrow p-n junction with a decreasing section on the forward branch of the volt-ampere characteristic (Fig. 1) and with a large conductance in the inverse direction. These properties are connected with the quantum mechanism of the penetration of the electrons through the potential barrier of the p-n junction via the tunnel effect. Among the favorable attributes of the tunnel diode is that it can be used at high frequencies and over a wide temperature range, its small dimensions, low sensitivity to nuclear radiation, and low energy consumption [6]. The limiting frequency of the tunnel diode is determined by the values of the parasitic capacitance C of the p-n junction, the differential resistance R at the operating point, and the loss resistance r (the equivalent circuit of the tunnel diode is shown in Fig. 2). Consequently the frequency properties of the tunnel diode are limited to a considerable degree only by the technological difficulties in manufacture, connected with the reduction in the aforementioned parameters. The conductivity mechanism itself should in principle be independent of the frequency up to 10^{13} cps [11].

Presently available tunnel diodes already have a limiting frequency reaching hundreds of gigacycles ($\lambda = 0.3$ cm) [12].

The foregoing properties of the tunnel diode permit its use in various radio circuits. From among the manifold applications, one of the most promising is the use of the tunnel diode as an amplifier for the UHF and the microwave bands. This is caused by the relatively low noise level (2-5 db), simple construction, and freedom from need of a pumping generator afforded by amplifiers using tunnel diodes.

In view of the considerable recent interest in tunnel-diode amplifiers, we make here an attempt to discuss systematically the problems connected with the operating principle of such systems, an evaluation of the results obtained, and the prospects of further developments.

The information cited in the article is based on data gleaned from published papers and also on research carried out by the authors.

1. AMPLIFYING PROPERTIES OF TUNNEL DIODE

Like any other element with a decreasing portion of the volt-ampere characteristic, the tunnel diode is capable of amplifying and generating electric oscillations.

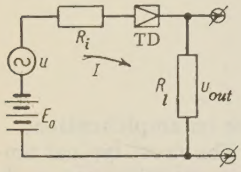


Fig. 3. Diagram of series-type amplifier (TD — tunnel diode)

The amplifying properties of such an element are best explained using as an example the operation of the simplest circuit illustrated in Fig. 3. We can write for this circuit the following obvious relation:

$$I(u_g + E_0) = I^2 R_l + I\varphi(I) + I^2 R_i,$$

where $V = \varphi(I)$ is the analytic expression for the volt-ampere characteristic of the tunnel diode.

In the general case $\varphi(I)$ is a nonlinear function. However, in amplification of alternating signals of sufficiently low amplitude the function $\varphi(I)$ can be represented as linear in the vicinity of the working points, the coordinates of which are I_0 and V_0 (see Fig. 1). Since the total current in the circuit I consists of a dc component I_0 and an ac component i, the equation for the energy balance in the circuit of Fig. 3, after averaging over the cycle, assumes the form

$$I_0 E_0 + \overline{iu_g} = I_0^2 R_l + \overline{i^2} R_l + I_0\varphi(I_0) + \overline{i^2}\varphi'(I_0) + I_0^2 R_i + \overline{i^2} R_i,$$

where $\varphi'(I_0)$ is the differential resistance of the tunnel diode at the operating point; $I_0\varphi(I_0) + i^2\varphi'(I_0)$ is the total power dissipated in the tunnel diode.

Inasmuch as in the linear approximation the presence of the ac component of the signal cannot influence the value of the direct current I_0, the dc power $I_0\varphi(I_0)$ dissipated in the tunnel diode remains constant during the process of operation. Therefore if the working point is chosen on the decreasing portion of the volt-ampere characteristic (see Fig. 1), $\varphi'(I_0) < 0$, then in the presence of an ac component the total power dissipated in the diode will decrease by an amount $\overline{i^2}\varphi'(I_0)$. This power is naturally divided between the load and the generator.

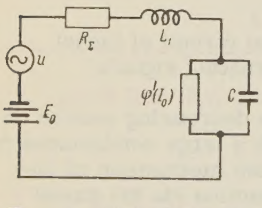


Fig. 4. High-frequency equivalent circuit of series-type amplifier.

Calculations show how an element with negative differential resistance can convert the energy of a dc source into an ac signal energy.

The circuit considered (Fig. 3) is the simplest tunnel-diode amplifier. As already shown, amplification with this circuit is possible only if the operating point is on the decreasing portion of the characteristic. This point will be stable only under certain relationships between the parameters of the equivalent circuit of the amplifier, as shown in Fig. 4. The circuit of Fig. 4 is obtained from that of Fig. 3 by replacing the tunnel diode with its equivalent circuit in accordance with Fig. 2. The resistance R_Σ and the inductance L_1 can correspond to the resultant active resistance and the summary inductance of the tunnel diode and the circuit.

The stability of circuits containing elements with dropping portions of the characteristic have been investigated in detail in [13]. The stability of tunnel-diode amplifiers has been treated in many papers [14, 15], and the gist of their conclusions is as follows.

The starting point for the determination of the stability condition is the characteristic equation, which for the equivalent amplifier circuit shown in Fig. 4 has the following form:

$$L_1 C\lambda^2 + \left[R_\Sigma C + \frac{L_1}{\varphi'(I_0)}\right]\lambda + \left[1 + \frac{R_\Sigma}{\varphi'(I_0)}\right] = 0,$$

where

$$R_\Sigma = R_i + R_l + r.$$

According to the Routh-Hurwitz criterion, a circuit described by a second-order differential equations with real coefficients will be stable, if all the coefficients of the characteristic equation have like signs. Consequently, if the operating point is chosen on the dropping portion of the characteristic, i.e., $\varphi'(I_0) < 0$ and $R_{\Sigma 1} > |\varphi'(I_0)|$, then $1 + \frac{R_{\Sigma 1}}{\varphi'(I_0)} > 0$ and the equilibrium position (point 2 on Fig. 1) is always unstable, since $L_1 C$ is a positive quantity. Therefore the case of Fig. 1 corresponds to the straight line $V = E_0 - 1R_{\Sigma 1}$. Inasmuch as $1\varphi'(I) > 0$ at points 1 and 3 of Fig. 1, these equilibrium positions are always stable and the circuit will be in one of these positions. A change of the circuit from one equilibrium position to the other can occur only under the influence of an external signal. As is well known, such a mode is realized in pulsed circuits.

Another operating condition, which must be satisfied in amplifier circuits corresponds to the case $\varphi'(I_0) < 0$ and $R_{\Sigma 2} < |\varphi'(I_0)|$. Then

$$1 + \frac{R_{\Sigma 2}}{\varphi'(I)} < 0$$

and the equilibrium position at the point I_0, V_0 will be satisfied only if

$$R_{\Sigma 2} C + \frac{L_1}{\varphi'(I_0)} < 0, \text{ i.e. } L_1 < R_{\Sigma 2} |\varphi'(I_0)| C.$$

If the inductance L_1 does not satisfy the last inequality, the circuit acts as a generator of undamped oscillations.

Thus, in choosing the parameters for tunnel-diode amplifier circuits it is necessary to satisfy two conditions

$$L_1 < R_\Sigma |R| C, \tag{1}$$

$$R_\Sigma < |R|,$$

where

$$R = \varphi'(I_0).$$

Depending on the type of amplifier circuit and on the way the load is connected, these conditions may change appreciably, but the characteristic feature is the fact that the equivalent resistance (load and losses), evaluated in parallel with the tunnel diode, should always be less than the absolute value of the differential resistance of the tunnel diode at the operating point.

In real tunnel-diode amplifiers, the ac and dc circuits are as a rule separated. In spite of this, the conditions for the stability should be satisfied to an equal extent in all the circuits. This requirement imposes certain specific conditions on the design of the bias circuits.

Methods for setting the working point in tunnel-diode amplifiers are quite varied. The choice of the bias source and the method whereby it is connected depend on many conditions such as the type of the circuit, the working frequency range, the differential resistance at the working point, and others. Within the framework of our review, we cannot dwell in detail on all the features involved in the solution of this problem. Various circuits for connecting the tunnel diode to the dc bias source are considered in detail in [16].

Principal tunnel-diode amplifier circuits

Inasmuch as the tunnel diode is a one-port network, the determination of the power gain of such amplifiers becomes particularly important.

Usually the power gain is defined in radio as the ratio of the signal power delivered to the load to the signal power available from the generator at the input to the given amplifier element. This is the so-called transducer power gain. Since the input and output of the tunnel diode cannot be distinguished, this definition of the power gain cannot be used for tunnel-diode amplifiers.

In addition to the transducer gain, one frequently uses the so-called nominal gain, defined as the ratio of the signal power delivered to the amplifier load to the power that the same signal generator can deliver to a matched load.

In any specific circuit, the input impedance of an amplifier element may not be matched to the internal impedance of the generator, and consequently the nominal gain characterizes the amplifying properties of the circuit as a whole (generator, amplifier, load) with account of the mismatch losses. Since such a definition of the gain does not require knowledge of the power at the input of the active element, it is convenient also for use with tunnel-diode amplifiers.

In addition to the nominal gain, amplifiers are frequently described in terms of a coefficient defined as the ratio of the signal power delivered to the load with the tunnel diode connected, to the power in the same load after the tunnel diode is disconnected. This coefficient is called the insertion power gain [14].

The inserted gain characterizes the power in the circuit when a tunnel diode is inserted in it, and this accounts for its name. The inserted gain differs from the nominal gain only in that in the definition of the nominal gain the input power is taken to be $u_g^2/4R_i$, whereas in the definition of the insertion gain we use

$$\left(\frac{u_r}{R_i + R_l}\right)^2 R_l.$$

Consequently, the nominal gain K differs from the inserted gain K_{ins} by a factor

$$\frac{K}{K_{ins}} = \frac{4R_iR_l}{(R_i+R_l)^2} = N \text{ omit.}$$

When $R_i = R_l$ we have $N = 1$, and in all other cases $N < 1$. Thus, for known R_i and R_l we can determine uniquely K from the known K_{ins} and vice versa. Both definitions of gain are used in the literature on tunnel diodes. Thus, for example, the nominal gain is used in [17] and the inserted gain in [14].

In the amplifier circuit, the tunnel diode can be connected either parallel to the load or in series with it [6]. Accordingly, there are two types of tunnel-diode amplifiers, parallel and series. The choice of the particular circuit depends on the frequency and amplification properties, and also on the structural features, which begin to play an appreciable role in the design of amplifiers for the UHF and microwave bands.

Parallel amplifier circuit

It is easy to show that for the parallel circuit (Fig. 5) the inserted gain is

$$K_{\text{в}} = \left| \frac{\frac{1}{R_p}}{\frac{1}{R_p} + \frac{1}{Z_d}} \right|^2, \tag{2}$$

where

$$\frac{1}{R_p} = \frac{1}{R_i} + \frac{1}{R_p},$$

and

$$Z_d = r - \frac{R}{1+(\omega RC)^2} + j\left[\omega L - \frac{\omega R^2 C}{1+(\omega RC)^2}\right], \tag{3}$$

Z_d is the tunnel diode impedance (in accordance with the circuit of Fig. 2). If we substitute (3) in (2) we obtain

$$K_{ins} = \frac{1}{1 + \dfrac{R_p^2[1+(\omega RC)^2] + 2R_p\{[1+(\omega RC)^2]r - R\}}{rR\left(\dfrac{\omega}{\omega_d'}\right)^2\left[\left(\dfrac{\omega_d'}{\omega_0}\right)^2 - 1\right]^2 + \left[r - R\left(1 - \left(\dfrac{\omega}{\omega_0}\right)\right)^2\right]^2}}, \tag{4}$$

where

$$\omega'_d = \frac{1}{RC}\sqrt{\frac{R}{r}}; \quad \omega_0^2 = \frac{1}{LC}.$$

When $r \ll R$, ω'_d will coincide with the frequency $\omega_d = \frac{1}{RC}\sqrt{\frac{R}{r} - 1}$ determined from the condition $\mathrm{Re}Z_d(\omega) = 0$. ω_d is called the limiting frequency of the tunnel diode [6], since at higher frequencies the real part of the tunnel-diode impedance becomes positive, i.e., the diode is no longer an element with negative resistance.

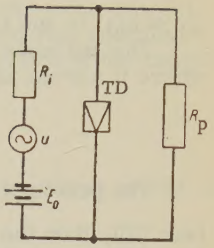


Fig. 5. Diagram of parallel-type amplifier.

According to (4) at frequencies $\omega \ll \omega_0$, recognizing that $r \ll R$, we obtain

$$K_{ins} = \frac{1}{\frac{1}{K_{inso}} + \frac{R_p^2}{rR}\left(\frac{\omega}{\omega_d}\right)^2}, \tag{5}$$

where

$$K_{inso} = \frac{1}{\left(1 - \frac{R_p}{R - r}\right)^2}. \tag{6}$$

An important parameter of the amplifier is the bandwidth, defined in terms of the specified coefficient of non-uniformity of the frequency characteristic $\gamma = K_{ins}/K_{ins0}$. It follows from (5) and (6) that the frequency ω_0, corresponding to a definite value of γ is

$$\omega_\gamma = \frac{\sqrt{\frac{1}{\gamma} - 1}}{\sqrt{K_{inso}} R_n C}.$$

According to (6) we have

$$R_n = \frac{\sqrt{K_{inso}} - 1}{\sqrt{K_{inso}}} R; \tag{6a}$$

Then

$$\omega_\gamma = \frac{\sqrt{\frac{1}{\gamma} - 1}}{\sqrt{K_{inso}} - 1} \frac{1}{RC}.$$

In particular, when $\gamma = 1/2$

$$\omega_\gamma (\sqrt{K_{inso}} - 1) = \frac{1}{RC}.$$

It is interesting to note that although the tunnel diode remains an active element up to a frequency ω_d, we have $K_{ins} = 1$ at a frequency

$$\omega_1 = \frac{\sqrt{K_{inso}} - 1}{\sqrt{K_{inso}} - 1} \frac{1}{RC} < \omega_d.$$

When

$$K_{inso} \gg 1 \quad \omega_1 = \frac{1}{RC}.$$

The limitation of the frequency range of the broad band amplifier to the frequency

$\omega_1 \ll \omega_d$ is due to the shunting action of the p-n junction capacitance.

The influence of the tunnel-diode capacitance is greatly reduced in resonant amplifiers, where it serves as part of the resonant circuit.

Parallel resonant amplifier circuit

The principal diagram of the resonant amplifier is shown in Fig. 6. If $L \ll \frac{R^2C}{[1+(\omega RC)^2]}$ (see [3]), then the inductance of the diode can be neglected, for in this case the tunnel diode has only a capacitive reactance. In this case the admittance of the circuit to the right of points 1 and 2 of Fig. 6 is

$$\frac{1}{Z_{1,2}} = \frac{1}{j\omega L_c} + \frac{1}{r - \frac{R}{1 - j\omega RC}}.$$

At the resonant frequency ω_r, determined from the condition $\frac{1}{Z_{1,2}} = 0$, the admittance between points 1 and 2 will be

$$\frac{1}{Z_{1,2}} \simeq -\frac{1 - \left(\frac{\omega_r}{\omega_d}\right)^2}{R}.$$

Substituting this expression into (2) in place of $1/Z_d$, we obtain the gain of the resonant amplifier

$$K_{intr} = \left\{\frac{1}{1 - \frac{R_p}{R}\left[1 - \left(\frac{\omega_r}{\omega_d}\right)^2\right]}\right\}^2. \qquad (7)$$

We see therefore that by choosing the inductance L_c and the resistance R_p, we can obtain an amplifier whose gain is greater than unity at a frequency close to ω_d. It is easily seen that the increase in the gain at high frequencies was accompanied by a reduction in the bandwidth. It is shown in [18] that the equivalent Q of circuits similar to the one of Fig. 6 is

$$Q_{eq} = \frac{Q}{1 + \frac{Z_r}{R_{eq}}},$$

where Q pertains to the L_cC circuit with loss resistance r; Z_r is the resonant impedance of the circuit; R_{eq} is the total impedance connected in parallel to the resonant circuit L_cC.

Consequently the bandwidth is

$$\Delta\omega = \frac{1}{R_pC}\left\{1 - \frac{R_p}{R}\left[1 - \left(\frac{\omega_r}{\omega_d}\right)^2\right]\right\} = \frac{1}{R_pC}\frac{1}{\sqrt{K_{intr}}}.$$

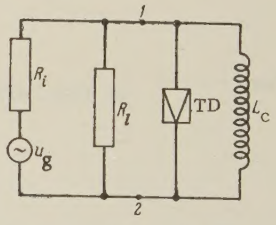


Fig. 6. Diagram of parallel resonant amplifier.

Using (7), we readily obtain

$$\Delta\omega\sqrt{K_{intr}} = \frac{1}{R_pC} = \frac{1 - \left(\frac{\omega_r}{\omega_d}\right)^2}{RC\left(1 - \frac{1}{\sqrt{K_{inso}}}\right)}. \qquad (8)$$

Thus, the product $\Delta\omega\sqrt{K_{intr}}$ is proportional to 1/RC (the same as for vacuum tubes) and decreases with increasing operating frequency ω_r. For most practical cases, when $K_{intr} \gg 1$ and $\omega_r \ll \omega_d$, relation (8) assumes the form

$$\Delta\omega\sqrt{K_{intr}} \simeq \frac{1}{RC}. \qquad (9)$$

This equation is used in many papers [11, 15] and is shown by calculations to be approximate. But since the assumptions made above as a rule hold true in real circuits, relation (9) is quite suitable for an estimate of the frequency and amplification properties of the tunnel diode.

Series amplifier circuit

According to the definition, the insertion gain for a series circuit such as shown in Fig. 3 has the following form

$$K'_{ins} = \left| \frac{R_i + R_l}{R_i + R_l + Z_d} \right|^2. \tag{10}$$

Substituting expression (3) for Z_d, we obtain

$$K'_{ins} = \frac{1 + \frac{R}{r}\left(\frac{\omega}{\omega_d}\right)^2}{\frac{1}{K'_{inso}} + \frac{R^2}{rR_\Sigma}\left(\frac{\omega}{\omega_d}\right)^2 + \left[\frac{R^2}{rR_\Sigma}\left(\frac{\omega}{\omega_d}\right)^2\right]^2}, \tag{11}$$

where

$$K'_{inso} = \frac{1}{\left(1 - \frac{R-r}{R_\Sigma}\right)^2}; \quad R_\Sigma = R_i + R_l.$$

It is easy to see that when

$$\frac{1}{K'_{inso}} \ll \frac{R^2}{rR_\Sigma}\left(\frac{\omega}{\omega_d}\right)^2 + \left[\frac{R}{rR_\Sigma}\left(\frac{\omega}{\omega_d}\right)^2\right]^2$$

and

$$\frac{1}{K'_{inso}} \simeq \frac{1}{K_{inso}} \ll 1, \text{ i.e. } R_\Sigma \simeq R \text{ and } R_p = R,$$

expression (11) coincides with (5). Thus, under the assumptions made above, broad band amplifiers have the same frequency characteristics whether they be series or parallel.

If we connect in series with the tunnel diode an inductance satisfying the resonance condition at the operating frequency ω_r (see (3))

$$\omega_r L_c - \frac{\omega_r R^2 C}{1 + (\omega_r RC)^2} = 0$$

and the stability condition (1), then the gain at the frequency ω_r will be

$$K'_{intr} = \left| \frac{1}{1 + \frac{1}{R_\Sigma}\left[r - \frac{R}{1 + (\omega_r RC)^2}\right]} \right|^2. \tag{12}$$

Thus, by choosing the inductance L_c and the resistance R_Σ, we can obtain considerable gain at close to the limiting frequency. It must be noted that the case $\omega_r \simeq \omega_d$ has no practical significance, since $R_\Sigma \to 0$ as $\omega_r \to \omega_d$. In practice a sufficiently large gain can be obtained only at frequencies which are five or ten times smaller than the limiting frequency of the tunnel diode.

The results of the calculations made for series and parallel amplifiers are illustrated by the plots of Fig. 7.

The relations given for the simplest typical cases suggest the following recommendations concerning the choice of amplifier circuits.

If the parameters of the tunnel diode (R, C, r, L) are known, together with the internal resistance R_i of the generator, the load resistance R_l, the working frequency ω_r, and the

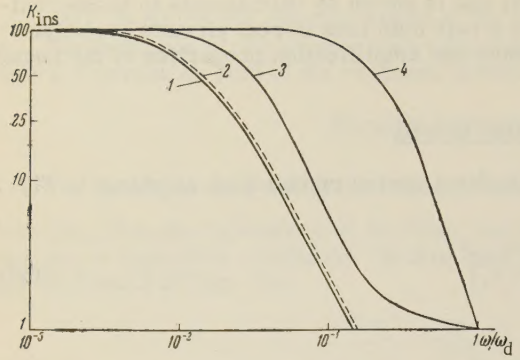


Fig. 7. Dependence of the gain on the frequency for various amplifier circuits: 1 — frequency characteristic of broadband parallel amplifier; 2 — frequency characteristic of broadband series amplifier; 3 — dependence of gain on the resonant frequency for series amplifier (see (12)); 4 — dependence of gain of parallel amplifier [see (7)].

bandwidth Δf, and if it is required to choose a circuit with a maximum gain, the main criterion for the choice of the circuit should be that the specified parameters satisfy the stability conditions for the operating point:

for series circuit $R_i + R_1 + r < |R|$ and $L < (R_i + R_1 + r)|R|C$;

for a parallel circuit $\frac{1}{R_i} + \frac{1}{R_1} > \frac{1}{|R - r|}$ and $L < \left(\frac{R_i R_1}{R_i + R_1} + r\right)|R|C$.

If all four inequalities are satisfied, then the choice of the circuit is determined by comparing (5) with (11) for broad band circuits and (7) with (12) for resonant circuits.

Brief analysis of microwave amplifier circuits

The calculations given above pertained to amplifiers with lumped constants. Greatest practical interest is attached to amplifiers in the microwave and UHF bands, in which an important role is played by the distributive character of the circuit elements.

Two types of microwave parallel amplifiers are possible, differing in the place where the tunnel diode is connected.

1) Transmission amplifier circuit, shown in Fig. 8, in which the tunnel diode is connected in the line between the load and the generator.

2) Reflecting circuit (Fig. 9), where the tunnel diode is connected after the load.

In the particular case when $Z_1 = Z_g = Z_0$, where Z_0 is the wave impedance of the line, the circuit of Fig. 8 will be stable if

$$\frac{1}{2}Z_0 + r < |R|$$

and

$$L < \left(\frac{1}{2}Z_0 + r\right)|R|C.$$

Calculations show that for a transmission circuit for all relationships between Z_0 and Z_1, and Z_g, the expression for the nominal gain has the following form

$$K = \frac{4}{\left|\left[Z_g\left(\frac{1}{Z_g} + \frac{1}{Z_p} + \frac{1}{Z_d}\right) + jZ_0\left(\frac{Z_r}{Z_0^2} + \frac{1}{Z_0} + \frac{1}{Z_d}\right)\operatorname{tg}\beta l_1\right]\cos\beta l_1\right|^2}.$$

If $Z_g = Z_0$, then

$$K = \frac{4}{\left|Z_0\left(\frac{2}{Z_0} + \frac{1}{Z_d}\right)\right|^2}.$$

For the frequency corresponding to $\tan\beta l_1 = 0$, where $\beta = 2\pi/\lambda$ is the wave number, we have

$$K = \frac{4}{\left|Z_g\left(\frac{1}{Z_g} + \frac{1}{Z_0} + \frac{1}{Z_d}\right)\right|^2}.$$

Analogous calculations for the reflecting circuit, shown in Fig. 9, yield the following

expression for the nominal gain:

$$K = \frac{4}{\left|\left[Z_g\left(\frac{1}{Z_g}+\frac{1}{Z_0}+\frac{1}{Z_1}\right)+i\left(1+\frac{Z_0}{Z_1}+\frac{Z_g}{Z_0}\right)\operatorname{tg}\beta l_1\right]\cos\beta l_1\right|^2},$$

where

$$Z_1 = Z_0\frac{Z_d + iZ_0\operatorname{tg}\beta l}{Z_0 + iZ_d\operatorname{tg}\beta l}.$$

If $Z_g = Z_0$, then

$$K = \frac{4}{\left|Z_0\left(\frac{2}{Z_0}+\frac{1}{Z_1}\right)\right|^2}.$$

These calculations show that in the particular case when $Z_g = Z_0$ and $Z_1 = Z_0$ the expression for the gain of the transmission amplifier coincides with the expression for the gain of the broad band amplifier with lumped constants (see (2)).

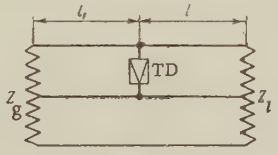


Fig. 8. Transmission microwave amplifier of the parallel type.

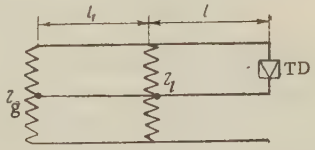


Fig. 9. Reflection microwave parallel amplifier.

For a reflecting type amplifier, this agreement will take place only subject to the additional condition $\tan\beta l = 0$ or $|Z_0| = |Z_d|$.

To increase the limiting amplification frequency, the capacitance of the tunnel diode can be compensated for with the aid of a short-circuited line segment (Fig. 10), the length of which should satisfy the condition

$$\operatorname{tg}\beta l_0 = -\frac{|Z_g|^2}{Z_0(\operatorname{Im} Z_g)}.$$

If we neglect the inductance and loss resistance of the tunnel diode, the last expression coincides with that obtained in [19].

The effect of a line segment of length l_0 in Fig. 10 is equivalent to connecting the conductance L_c in Fig. 6. Therefore the expression for the gain of a transmission amplifier with a short-circuited line segment coincides with $Z_g = Z_1 = Z_0$ with the expression (7) for the parallel resonant amplifier with lumped constants.

A typical series circuit comprising an amplifier with line is shown in Fig. 11.

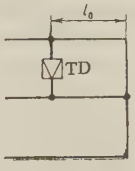


Fig. 10

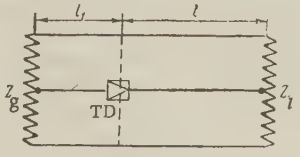


Fig. 11

Fig. 10. Diagram showing compensation of capacitance of tunnel diode at microwave frequencies.

Fig. 11. Diagram of microwave amplifier of the series type.

Analysis of this circuit leads to the following expression for the nominal gain (when $Z_1 = Z_0$):

$$K = \frac{4|Z_0|^2}{\left|\left[(Z_d + Z_0 + Z_g) + j\left(Z_0 + \frac{Z_g}{Z_0}(Z_d + Z_0)\right) \operatorname{tg} \beta l_1\right] \cos \beta l_1\right|^2}.$$

If $Z_g = Z_0$, then the gain is

$$K = \frac{4|Z_0|^2}{|Z_d + 2Z_0|^2}.$$

If $Z_g \neq Z_0$ but $\tan l_1 = 0$, then

$$K = \frac{4|Z_0|^2}{|Z_d + Z_0 + Z_g|^2}.$$

As can be readily verified, the last relation coincides in the case of a series circuit of a distributed-element amplifier with expression (10) for a series amplifier with lumped constants.

2. NONLINEAR PROPERTIES OF TUNNEL DIODES

All the relations given in the preceding sections for the calculation of the frequency and amplification properties of tunnel-diode amplifiers are valid, strictly speaking, only for signals with infinitesimally small amplitudes, when the characteristic of the tunnel diode in the vicinity of the working point can be represented by means of a linear function. As is well known, the real characteristic of the tunnel diode is essentially nonlinear, and the amplitudes of the amplified signals may exceed considerably the permissible limit up to which the volt-ampere characteristic can be regarded as linear [36]. In addition, the instability of the fixed bias source may also lead to a change in the position of the operating point, and consequently to a deviation of the real value of the gain from the value calculated by the linear theory.

To estimate the magnitude and character of the nonlinearity let us consider in greater detail the characteristic of the tunnel diode shown in Fig. 12a. The dashed lines in Fig. 12a are the individual components of the currents flowing through the tunnel diode in the forward direction. Curve 1 corresponds to the resultant tunnel current through the junction and has a bell-shaped form; curve 2 corresponds to the normal diffusion current; curve 3 represents the dependence of the excess current, connected with the presence of additional impurity levels in the forbidden band of the semiconductor [20, 21, 22] on the voltage on the tunnel diode. The dependence of the excess current on the voltage has not yet been sufficiently well studied, and curve 3 of Fig. 12a is only one of the possible approximations, adopted in Ref. [5].

As follows from Fig. 12a, the resultant curve (solid line) obtained as a result of adding all the foregoing current components has a drooping portion of the characteristic, located between a rather sharp maximum (I_{max}) and a fairly smeared minimum (I_{min}). Inasmuch as the operating point in amplifier circuits is chosen on the drooping portion of the characteristic, greatest interest attaches to the character of the relationship $R = f_1(V)$ (see Fig. 12b) or $G = f_2(V)$ (see Fig. 12c) in the voltage interval from V_1 to V_2.

Calculations based on the linear theory show that to ensure a specified gain the operating point can be chosen anyplace on the drooping portion; it is immaterial here what the value of $|R|$ is, since by choosing the corresponding load resistance one can always ensure as large a gain as is convenient. However, as follows from an analysis of the real characteristics of Figs. 12b and c, the most suitable is an operating point at a voltage V_0 corresponding to the minimum value of $|R|$ or the maximum transfer conductance $|G|$. This conclusion is the result of the fact that the differential values $|R|$ or $|G|$ change very little in the vicinity of the point I_0, V_0. Thus, at the point I_0, V_0, we can ensure minimum nonlinear distortion at maximum output power. In addition, in the case when the operating point is shifted relative to V_0, there exists a danger of self excitation of the circuit whenever the voltage on the tunnel diode changes, since one of the conditions of stability can be violated near V_0. The choice of bias at the point V_0 is also advantageous because in this case the limiting frequency ω_d and the

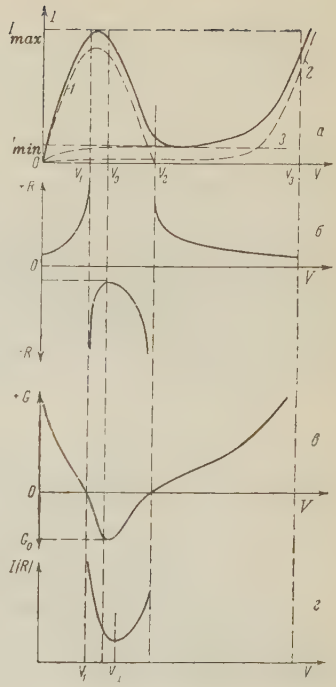


Figure 12

a — volt-ampere characteristic of tunnel diode: 1 — dependence of tunnel current on the voltage; 2 — dependence of diffusion current on the voltage; 3 — dependence of excess current on the voltage;
b — dependence of differential resistance of tunnel diode on the voltage;
c — dependence of differential conductance of the tunnel diode on the voltage;
d — dependence of the product $I|R|$ on the voltage.

product $\Delta f\sqrt{K}$ will have maximum values.

The choice of the operating point can be greatly influenced by the shot noise of the tunnel diode, which as shown in references [23, 24] is proportional to the product $I|R|$. As follows from Fig. 12d, this product has a minimum at a voltage $V = V_m$ which turns out to be shifted somewhat to the right relative to V_0. Therefore the final choice of the operating point on the volt-ampere characteristic depends on the purpose of the amplifier and is determined in each particular case on the basis of a compromise solution.

In the present time there is no simple analytic expression for the tunnel-diode volt-ampere characteristic [5, 25, 26, 27, 49], which would describe sufficiently accurately all the processes occurring in narrow p-n junctions formed by two degenerate semiconductors. All the presently known analytic expressions for the volt-ampere characteristic are either polynomials of not less than the eighth degree, or contain different combinations of exponential terms. Therefore to estimate the nonlinear properties of the tunnel diode we shall use the approximate expression for the conductivity G(V):

$$G(V) = \alpha (V - V_0)^2 - G_0, \tag{13}$$

where $\alpha = \text{const}$. This expression, as verified experimentally, approximates sufficiently accurately the dependence G(V) in a small vicinity ($\Delta V < 0.1\ V_0$) of the working point V_0.

In amplifying circuits the dependence of G on V causes the gain to become a function of the bias voltage and of the ac signal amplitude, i.e., $K = f(V + A)$, where A is the amplitude of the alternating signal, $V = V_0 + \delta V$ — the constant bias, and δV — fluctuation of the power supply voltage.

The relative change in the gain can be determined in the following fashion

$$\eta_V = \frac{\Delta K}{K_{inso}} = 1 - \frac{K_{ins}(V + A)}{K_{inso}}, \tag{14}$$

where $K_{ins0} = 1/(1 - R_p G_0)^2$ — value of the insertion gain at $A \to 0$ and $V = V_0$.

According to (2), the expression for $K_{ins}(V + A)$ can be represented as follows:

$$K_{ins} = \frac{1}{(1 - R_l G_{ef})^2}, \tag{15}$$

where G_{ef} is the effective transfer conductance of the characteristic at the operating point, averaged over the cycle.

The value of G_{ef} can be obtained from energy considerations. The mean energy delivered by the tunnel diode during one cycle to the load should be equal to the mean energy delivered to the effective conductance G_{ef} during the same time interval:

$$G_{ef} = G_0 - \alpha \frac{\int_0^{2\pi} (\Delta V)^4 d\varphi}{\int_0^{2\pi} (\Delta V)^2 d\varphi} = G_0 - \frac{3}{4}\alpha A^2 \frac{\frac{8}{3}\beta^4 + 4\beta^2 + 1}{2\beta^2 + 1}, \quad (16)$$

where $\varphi = \omega t$; $\beta = \delta V/A$; $\Delta V = A \cos \varphi + \delta V$.

Substituting (16) in (15) we obtain an expression for K_{ins}, which in general will not coincide with the value of the gain obtained from the linear theory.

One must note two particular cases, which can be encountered with equal probability in practice. The first corresponds to an amplification mode in which the amplitude of the alternating signal A at the diode is much less than the fluctuations δV, due to the instability of the power supply. Then $\beta \ll 1$ and the expression for G_{ef} assumes the form

$$G_{ef} = G_0\left(1 - \frac{3}{4}\frac{\alpha A^2}{G_0}\right). \quad (17)$$

Using (14), (15), and (17) we obtain for the relative change in the gain

$$\eta_y = 1 - \frac{1}{\left(1 + \frac{3}{4} R_p \alpha A^2 \sqrt{K_{inso}}\right)^2}.$$

For typical values of the parameters ($V_0 - V_1 \gtrsim 50$ mv, $R_p G_0 \sim 1$) and $A \leq 5$ mv, $\sqrt{K_{ins}} < 10^2$ the second term in the parentheses proves to be less than unity, and the last expression simplifies to

$$\eta_a \simeq \frac{3}{2} R_p G_0 \left(\frac{A}{V_0 - V_1}\right)^2 \sqrt{K_{inso}}$$

from which it follows that for given values of η_a the amplitudes of the signals A_1 and A_2, for two amplifiers with gains K_{ins01} and K_{ins02}, are connected by

$$\frac{A_1}{A_2} = \sqrt[4]{\frac{K_{inso2}}{K_{inso1}}} \frac{R_{p_2}}{R_{p1}}.$$

Thus, the nonlinearity of the volt-ampere characteristic imposes essential limitations on the permissible signal amplitude. In amplifiers using germanium tunnel diodes, even when the gain is one the order of 25 db, the maximum signal amplitude on the diode must not exceed an order of 2 mv. Since usually the load is connected in parallel to the tunnel diode and its resistance is approximately equal to the differential resistance of the diode at the operating point, the maximum value of the signal power delivered to the load is approximately $P = A^2/R$. If $A \simeq 2$ mv, then $P \simeq 4 \cdot 10^{-6}/R$ watts.

It follows therefore that, from the point of view of increasing the output power, it is advantageous to use tunnel diodes with small values of P. The use of such diodes, however, entails additional difficulties, connected with the reduction in the ohmic loss in the wires, reduction of the internal resistance of the power supply, and suppression of parasitic oscillation. Therefore for germanium tunnel diodes, in which $|R| = 10 - 150$ ohms, the output power usually does not exceed one microwatt. If we neglect the nonlinear distortion and assume that the possible peak to peak oscillations in the tunnel diode is given by the quantity $V_2 - V_1$ (see Fig. 12a), we can write for an estimate of the maximum output power of germanium tunnel diodes the following approximate relation [30]:

$$P_{max} = \frac{10^{-2}}{R} \text{ w}$$

The second particular case corresponds to a system intended for the amplification of very weak signals, when the amplitude A of the oscillations is much less than the fluctuations of the bias voltage δV. In this case $\beta \gg 1$ and according to (16) we have

$$G_{ef} = G_0 - \alpha (\delta V)^2.$$

Then the relative change in the gain, connected only with the instability of the power supply, will be

$$\eta_a = 1 - \frac{1}{[1 - \alpha (\delta V)^2 R_p \sqrt{K_{inso}}]^2} \simeq 2\alpha (\delta V)^2 R_p \sqrt{K_{inso}} = 2 \left(\frac{\delta V}{V_0 - V_1} \right)^2 \sqrt{K_{inso}}$$

hence

$$\eta_E = \frac{\delta V}{V_0} = \sqrt{\frac{\eta_a}{2 R_p G_0 \sqrt{K_{inso}}}} \left(1 - \frac{V_1}{V_0} \right).$$

The last equation enables us to calculate the permissible relative change in power voltage η_E for a specified coefficient η_a.

Fig. 13 shows the variation of η_E with η_a, plotted for two values $K_{inso} = 100$ and $K_{inso} = 900$, under the assumption that $R_p G_0 = 1$ and $V_1/V_0 \simeq 5/8$.

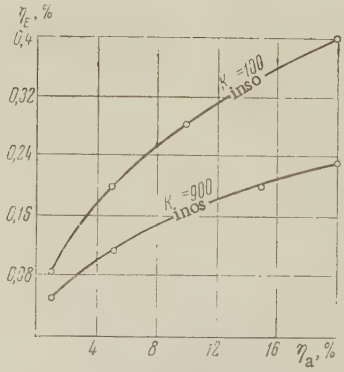


Fig. 13. Dependence of the instability coefficient of the amplifier η_a on the instability coefficient of the power supply η_E.

As follows from Fig. 13, the requirements imposed on the stability of the power sources are quite high. This circumstance may be very important in the construction of stable amplifiers with high gain.

Temperature dependence of the gain

When speaking of stability of amplifiers, we must discuss in particular the temperature dependence of the parameters of tunnel diodes, which determines to a greater degree the reliability of operation of the entire circuit as a whole.

In spite of the fact that the tunnel effect in narrow p-n junctions is observed over a wide range of temperatures (from -269 to 200 or 300°C) [6], the variation of the parameters of real diodes is quite significant even in the interval ±50°C. Theoretical and experimental researches, devoted to the temperature dependence of tunnel diodes, carried out by various authors [21, 28], show that the most affected is the excess current, which determines the form of the characteristic in the region of the minimum. Less strongly varying is the maximum current. In this case the direction of its change depends to a very great degree on the type of the semiconductor and on the degree of doping [22, 28, 29]. This complicated character of the temperature dependence is connected with a large number of physical parameters (position of the Fermi level, width of the forbidden band, effective mass, etc.), which vary differently with the temperature, depending on the type of material and the concentration of the impurity added. The resultant process can therefore have temperature coefficients of either sign. Usually germanium diodes doped with indium, gallium, and arsenic with specific resistivity on the order of 6×10^{-4} ohm-cm, display a decrease in I_{max} and an increase in I_{min} with increasing temperature. As a result, the modulus of the differential resistance increases with increasing temperature.

As shown in reference [19], devoted to an investigation of the temperature stability of an amplifier, the best effect is obtained by stabilizing the current at the operating point, and the worst is obtained by stabilizing the voltage. From the physical point of view this result can be explained by the fact that in an ideal tunnel diode ($I_{min} = 0$), the transfer conductance of the characteristic, on which the gain depends, is proportional, to the probability of tunneling through the junction. Inasmuch as the current through the junction is also proportional to the tunneling probability, by stabilizing the current at the operating point we can expect that we thereby stabilize also the gain. The correctness of stabilization of this type was qualitatively confirmed in experiments carried out by the authors.

Figure 14 shows the experimental temperature dependence of the change in voltage V_0,

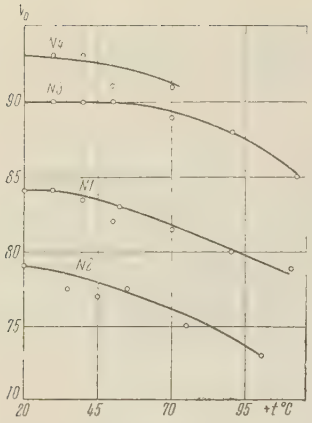


Fig. 14. Dependence of the position of the point of the inflection on the temperature for four tunnel diodes.

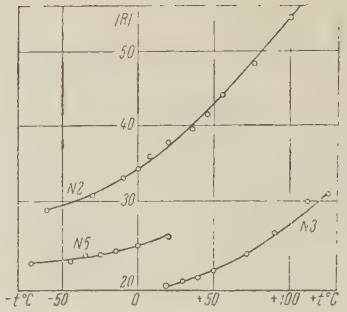


Fig. 15. Dependence of the modulus of the differential resistance on the temperature for three tunnel diodes.

corresponding to the point with maximum transfer conductance. As can be seen from Fig. 14, the point of inflection on the volt-ampere characteristic shifts to the left towards the origin for all tested diodes. This tendency will take place also if the current stabilization of the operating point is effected in the circuit.

However, investigations of families of the temperature characteristics of tunnel diodes show that it is impossible to obtain sufficiently good results over a wide temperature interval by stabilizing the current at the operating point. This apparently is connected with the fact that in real diodes the form of the volt-ampere characteristic is very greatly influenced by the excess current, which has a different temperature coefficient [21] than the maximum current. The greater the range of variation of the temperature, the stronger the influence of the excess current, and the transfer conductance at the operating point begins to depend more strongly on the temperature.

To illustrate the quantitative change of the modulus of the differential resistance with temperature, at the point of inflection of the characteristic V_0, Fig. 15 shows several experimental curves for germanium diodes with specific resistivity $\rho \simeq 6 \cdot 10^{-4}$ ohm-cm. Even a superficial study of the curves of Fig. 15 shows that temperature stabilization of tunnel-diode circuits is a rather pressing problem.

3. NOISE PROPERTIES OF TUNNEL-DIODE AMPLIFIERS

One of the most important characteristics of the tunnel-diode amplifier is the noise that it introduces. The sources of noise in tunnel-diode amplifiers are: 1) shot noise, due to the current flowing through the p-n junction, 2) low-frequency noise, due to leakage through the junction, the spectral density of which decreases as 1/f, 3) thermal noise in the loss resistance of the tunnel diode, 4) thermal noise in the load [30].

As shown in [30], the 1/f noise in good tunnel diodes is much less than the shot noise even at frequencies above 1 kc, and is therefore usually neglected in the calculation of the noise properties of tunnel-diode amplifiers.

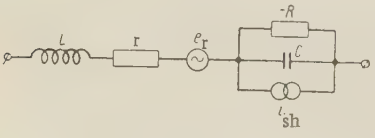


Fig. 16. Equivalent noise circuit of tunnel diode.

The equivalent noise circuit of a tunnel diode is shown in Fig. 16 [31, 32], where the following symbols are used: e_r — noise voltage of the loss resistance r, i_{sh} — shot noise through the p-n junction. The mean-squared values of e_r and i_{sh} are given by the expressions

$$\overline{e_r^2} = 4kT_r r \Delta f,$$

$$\overline{i_{sh}^2} = 2q\,(\gamma^2 I_{c\to v} + I_{v\to c})\Delta f,$$

where k — Boltzmann's constant; T_r — loss-resistance temperature; q — electron charge; Δf — bandwidth; $I_{c\to v}$ — mean current due to the tunneling of the electrons from the conduction band of n region into the p region; $I_{v\to c}$ — mean current due to the tunneling of the valence electrons from the p region to the n region; γ^2 — "smoothing" coefficient [33].

The "smoothing" of the current $I_{c\to v}$ is due to the greater probability of transition of the electrons from the conduction band of the n region into the p region. Only the current $I_{c\to v}$ can be smoothed, since the transition probability of the valence electrons from the p region is small. The coefficient γ depends on the structure of the tunnel diode.

It is shown in [23] that at sufficiently large biases $V > V_1$ (see Fig. 12a) we have $I_{v\to c} \ll I_{c\to v}$. In this case the resultant current through the tunnel diode, the value of which is $I_{c\to v} - I_{v\to c} = I_0$, practically coincides with the current $I_{c\to v}$ and the mean square shot current is found to be $\overline{i^2}_{sh} = 2q\gamma^2 = 2q\gamma^2 I_0$.

An investigation of noise in germanium tunnel diodes at frequencies 0.5 and 100 megacycles [34] has shown that at a bias voltage greater than 4 kt/q the noise is determined essentially by the shot processes, and at lower voltages they approach the level of thermal noise.

One of the quantitative characteristics of the noise properties of an amplifier is the noise factor, a quantity defined as the ratio of the noise power at the output of a real amplifier to the noise power at the output of an ideal amplifier, i.e., which does not add noise of its own, and which operates with the same signal generator. Using the equivalent noise circuit of the tunnel diode (see Fig. 16) we can calculate the noise factor of a specific amplifier.

The calculations show that the noise factor of a broadband parallel amplifier (Fig. 5) at frequencies $\omega \ll \omega_d$ is determined by the expression

$$F = 1 + \frac{T_l}{T_g}\frac{R_i}{R_l} + \frac{qI_0}{2kT_g}R_i + \frac{T_r}{T_g}\frac{rR_i}{R^2}, \tag{18}$$

and at frequencies $\omega \sim \omega_d$

$$F = 1 + \frac{T_l}{T_g}\frac{R_i}{R_l} + \frac{qI_0}{2kT_g}R_i + \frac{T_r}{T_g}\frac{R_i}{2R}, \tag{19}$$

where T_l and T_g are the load and generator temperatures.

If we neglect in (18) the noise in the loss resistance r, this expression agrees with the equation obtained in [35].

It is easy to show that at small values of loss resistance r, the expression for the noise factor of a parallel resonant amplifier (Fig. 6) coincides with expression (18). Indeed, since at the resonant frequency the tank-circuit resistance is purely active, the equivalent circuit of the resonant amplifier will be analogous to the low-frequency circuit of a broadband amplifier. The only difference will be that in the resonant amplifier we have in parallel with the negative resistance of the tunnel diode, not only the load and generator resistances, but also the resonant circuit of the tank circuit.

The expression for the noise factor of a resonant amplifier with r = 0 is given in [17]. It coincides with expression (18), except for the symbols.

As can be seen from (18) and (19), the term $T_l R_i / T_g R_l$ can be greatly reduced by reducing the ratio R_i/R_l. If we take $R_i/R_l \ll 1$, it follows from (6a) and (7) with sufficiently large gains at the frequencies $\omega \ll \omega_d$, that $R_i \cong R$. In this case, inasmuch as $r \ll R$ always, we obtain

$$F \simeq 1 + \frac{qI_0R}{2kT_g}. \tag{20}$$

For a resonant amplifier, inasmuch as ω_r can be close to ω_d in accordance with (7), we obtain for $K_{intr} \gg 1$

$$R_i \simeq \frac{R}{1 - \left(\frac{\omega_r}{\omega_d}\right)^2},$$

Hence, under the assumptions made above

$$F \simeq 1 + \frac{qI_0}{2kT_r} \frac{R}{\left[1 - \left(\frac{\omega_r}{\omega_d}\right)^2\right]}.$$

Thus, when $\omega_r \to \omega_d$, the noise factor of the resonant amplifier increases without limit. The same result was obtained in [31, 32].

In a broadband amplifier the condition $R_p < R$ is always satisfied. Therefore the noise factor remains a bounded quantity in this case at all frequencies.

For a series resonant amplifier we can show that its noise factor in this case of sufficiently large gain is given by the expression

$$F = 1 + \frac{T_l}{T_g} \frac{R_l}{R_i} + \frac{qI_0}{2kT_g} \left(1 + \frac{R_l}{R_i} + \frac{r}{R_i}\right) R + \frac{T_r}{T_g} \frac{r}{R_i}. \tag{21}$$

Unlike (18), the noise factor for the series circuit increases with decreasing ratio R_i/R_l. If we take $R_i/R_l \ll 1$, then at frequencies $\omega \ll \omega_d$ we get $R_i \simeq R$, and we obtain for the noise factor at low frequencies

$$F \simeq 1 + \frac{qI_0R}{2kT_g}. \tag{22}$$

Under the assumptions made, as follows from a comparison of (20) with (22), the expressions for the series and parallel circuits are the same.

In the general case, for the series circuit, as can be seen from expression (12), the following equation is valid for large gain

$$R_\Sigma = R_i + R_l \simeq \frac{R}{1 + (\omega RC)^2} - r.$$

If $R_i/R_l \gg 1$, we can neglect in this last expression the load resistance R_l. If the resultant value of R_i substituted in (21), the expression for the noise factor coincides with the expression given in [32]. It follows from the last relations that the noise factor increases sharply in a series resonant amplifier at frequencies close to the limiting value.

The noise factor is not a sufficient criterion for the applicability of an amplifier stage, since it does not characterize its amplification properties. Therefore for a more complete description of the amplifier one frequently introduces the quantity M, defined in the following fashion [37, 40]:

$$M = \frac{F - 1}{1 - \frac{1}{K}},$$

where K — gain of the stage. The last relation shows that the amplifier should be so designed as to make M minimal.

Using (20) and (5) we obtain for a parallel broad band amplifier with $R_i/R_l \ll 1$

$$M \simeq \frac{qI_0R}{2kT_g[1 - (\omega RC)^2]}.$$

At the frequencies $\omega \to 1/RC$ the quantity M increases without limit. This shows that the use of the stage of frequencies close to limiting is meaningless in practice.

Under the same assumptions ($R_i/R_l \ll 1$ and $K_{ins\,0} \gg 1$) we obtain for the parallel resonant amplifier

$$M \simeq \frac{qI_0R}{2kT_g\left[1 - \left(\frac{\omega_r}{\omega_d}\right)^4\right]},$$

from which it follows that $M \to \infty$ when $\omega_r \to \omega_d$. Therefore the resonant amplifier also is

useless at frequencies close to the limiting frequency of the tunnel diode.

In all the approximations given above, the term characterizing the shot noise is proportional to the product I_0R. The value of I_0R depends on the choice of the operating point.

A theoretical calculation of the bias voltage V_m at which the product I_0R has a minimum value is given in [38, 39], from which follow purely qualitative conclusions regarding the form that the volt-ampere characteristic of the tunnel diode with the best noise properties should have. Experiments carried out by the authors show that the voltage V_m is shifted to the right relative to V_0 by 10-20 mv. The measurements were carried out for germanium tunnel diodes.

At the point V_0 the value of I_0R is greater than the minimum value by approximately 5-10 percent. Therefore if such an increase in the shot noise is not acceptable, the operating point must be chosen to the right of V_0. This yields a certain gain in the noise, but, as was shown earlier, it greatly deteriorates the linearity and stability of the amplifier.

The minimum value of the noise factor of amplifiers with tunnel diodes is estimated to be 2-3 db at room temperature. The better experimental samples of amplifiers have at present noise factors on the order of 2 db [41].

4. CONSTRUCTION OF TUNNEL-DIODE AMPLIFIERS

The most obvious advantages of the tunnel diode above the presently known semiconductor and vacuum tube devices is the low energy consumption (10^{-3} — 10^{-4} watt), high frequency limit of amplification ($\sim$ 100 gigacycles), the possibility of producing miniaturized apparatus, and the relatively low noise level.

Unfortunately the first advantage involving the low energy consumption, can at the present time still not be utilized, since to produce a fixed bias one usually employs standard cells with voltage 1.5 - 2.2 v and the excess voltage is consumed in the ballast resistor.

As a result, the power drawn from the power supply is as a rule greater than or equal to the power necessary to feed analogous transistor circuits.

Since from the point of view of the dimensions the tunnel-diode circuits do not have noticeable advantages over transistor circuits, one can say that at frequencies up to about 100 megacycles, where modern transistors have sufficiently good amplification properties, the use of tunnel-diode circuits is hardly justified. Another fact in favor of this conclusion is that cascade connection of tunnel-diode amplifiers involves a directivity problem, the solution of which entails great difficulties in the indicated frequency band. On the other hand, the construction of multi-stage transistor amplifiers at frequencies up to 10^8 cps does not cause serious difficulties.

Perhaps the only advantage of tunnel-diode amplifiers in this frequency band is their lower noise level.

At these frequencies it is also possible to use combined circuits in which tunnel diodes are used matching interstage elements in transistor amplifiers [42].

Thus, the use of tunnel-diode amplifiers is most promising in the UHF and the microwave bands, where their advantages are most clearly pronounced.

The main difficulties arising in the construction of tunnel-diode amplifiers in the indicated frequency bands are connected essentially with questions of matching the differential resistance of the tunnel diode with the corresponding distribution systems.

Fig. 17 shows a specific resonant amplifier [14] using a series circuit, operating at 100 megacycles with a gain of 32 db. The amplifier employs a germanium diode with parameters C = 5 micromicrofarad, R = 143 ohms, r = 2 ohms. All the values of the parameters of the circuit are uniquely determined by relations (1) and (12). The generator and the load are connected to the circuit with the aid of 50-ohm lines, matched at the terminals. The matching is a necessary condition for normal operation of the amplifier, for in this case no reactive parameters are introduced from the generator and load sides. The inductance L_1 and the capacitance of the diode form a resonant circuit tuned to 100 megacycles. The inductance L_2 acts as a choke, decoupling the fixed bias circuit from the amplifier circuit. Ballast resistances R_1 and R_2 are chosen such as to ensure on the tunnel diode a voltage drop of 125 millivolts, which is essential to set the operating point.

A typical construction of a resonant amplifier, described in [19] and made up into a parallel circuit and operating in the UHF band is shown in Fig. 18. The tunnel diode is connected in a strip-type cavity, from which the leads to the generator and the load are in the form of coaxial lines. The construction provides for a matching rod, so as to insure normal

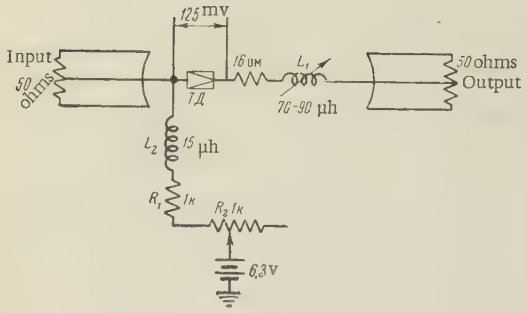


Fig. 17. Diagram of series resonant amplifier (f = 100 Mc).

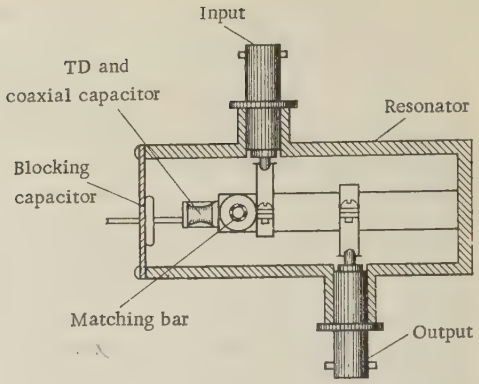


Fig. 18. Construction of parallel type amplifier (f = 460 Mc).

operation of the amplifier as the SWR in the antenna changes from 1 to 1.8 db. The resonator is tuned with the aid of a coaxial capacitor bar, connected in parallel to the tunnel diode. To reduce the reactive components of the inductance of the leads to the tunnel diode and to the power supply, a blocking capacitor of 1000 micromicrofarad is connected at the end of the central conductor. This amplifier operates in a range from 405 to 460 Mc with a gain of 15 db, a noise factor of 5.5 db, and a bandwidth of 12 mc. The tunnel diode used in the circuit had the following parameters: C = 7 micromicrofarad, R = 100 ohms, r = 1 ohm, and lead inductance 6×10^{-9} henry. The dimensions of the amplifiers together with power supply were 48 x 29 x 16 mm. The weight of the amplifier is on the order of 100 grams.

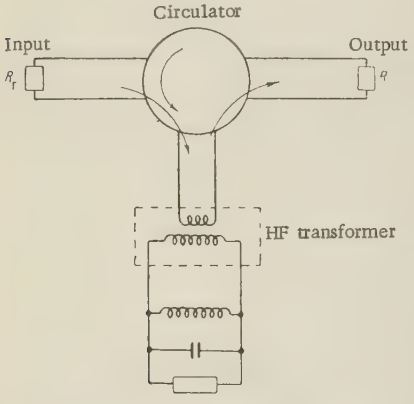


Fig. 19. Block diagram of amplifier with circulator.

To broaden the bandwidth of tunnel-diode amplifiers frequently band filters are used instead of tuning the resonant circuit to a fixed frequency [43]. In this case it is possible to increase appreciably the product $\Delta f \sqrt{K}$. Amplifiers of this type are described in detail in [44]. Structurally the filter is made in the form of strip lines. The amplifier operates at 330 mc, the insertion gain is 25 db, the bandwidth is 10 mc, and the calculated noise factor is 2.7. The parameters of the tunnel diode are: R = 133 ohms, C = 5 micromicrofarad, r = 1 ohm, $L = 10^{-9}$ henry.

As already noted, the main shortcoming of tunnel diodes is that they are non-directional. This makes it quite difficult to construct multi-stage amplifiers. These difficulties are overcome by two methods [15] extensively used in microwave technology.

The first calls for a circulator to be used in the circuit (Fig. 19). The signal from the generator is fed to the amplifier through a ferrite coupler (circulator). The amplified signal is reflected to the circulator, which directs it only towards the load. Thus, the transmission of the amplified signal is only in one direction. The use of a circulator in the circuit reduces the noise factor and ensures sufficiently stable operation of several stages of tunnel-diode amplifiers.

The practical circuit and the construction of the tunnel-diode amplifier with circulator are described in [45]. This amplifier operated at 4.5 gigacycles with a gain of 23 db, bandwidth 20 mc, effective noise temperature 1200° K (7 db), and maximum output power 10^{-6} w.

The use of circulators is most advantageous in the centimeter band, where the dimensions can be sufficiently small. In the UHF band the dimensions of ferrite couplers may prove to be excessively large and this, as a rule, limits their applicability at such frequencies.

Another method permitting directional transmission of the signal is to use hybrid waveguide junctions [41, 46]. A typical block diagram of an amplifier using such a junction is

shown in Fig. 20. The main element of this circuit is a hybrid ring, made in the form of a coaxial line or waveguide. As shown in Fig. 20, the generator, the load, and the two tunnel-diode amplifiers are connected to different points of the hybrid ring.

Inasmuch as the length of the line between the second amplifier and the rings is one quarter of a wavelength greater than l, then in accordance with the designations of Fig. 20, the phase relationships in the circuit cause the signals reflected from the amplifiers to arrive at the generator in phase opposition and do not act on it.

On the other hand, because of the same phase relations in the hybrid ring, these signals will add up at the load. As a result, the amplifier will have directivity. Unlike circuits with circulators, amplifiers using the hybrid circuit are quite sensitive to a mismatch in the load and generator circuits. Therefore for them to operate stably it is necessary as a rule to use additional matching elements.

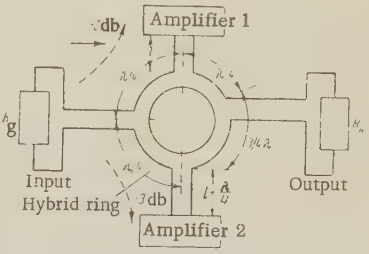


Fig. 20. Block diagram of amplifier with hybrid ring

Theoretical and experimental investigations of the amplifier with hybrid junction are described in [41]. The amplifier operated stably in the frequency band 210 to 625 mc with a gain of 8.2 ± 0.6 db. The noise factor at 350 mc was 0.93 ± 0.4 db. Normal operation of the amplifier was not disturbed at a SWR ≤ 2 db at the input and ≤ 1.05 at the output. The loss of gain in the hybrid junction between the generator and the amplifiers was 3 db. A tunnel diode j56z was used in the amplifier (I_{max} = 11.1 ma, R = 115 ohms, f_d = 2.1 gigacycles).

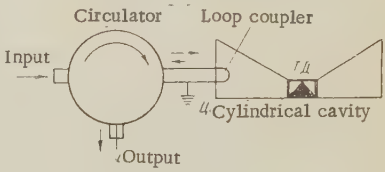


Fig. 21. Construction of microwave amplifier with cylindrical cavity and circulator

f, Gc	K, db	F, db
6.8	25	7.5
9.7	22	10.7
10.8	26	11.5
25.8	36	—

From the point of view of maximum amplified frequency, which can be obtained with modern tunnel diodes, great interest attaches to reference [47], which reports results of an investigation of several amplifier circuits using gallium-arsenide tunnel diodes.

The tested diode was placed in the center of a cylindrical cavity, which was conical in cross section (Fig. 21). The resonator operated in reflection and therefore had only a single input in the form of a coupling loop, through which the signal was fed from the generator and the amplified reflected signal was fed to the load. The separation between the input signal and the output signal was with the aid of the circulator.

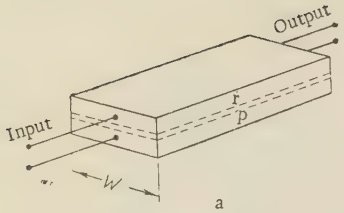
The circuit of Fig. 21 [47] was used to investigate amplifiers operating at frequencies 6.8, 9.7, 10.8, and 25.8 gigacycles. The main results of the measurements are listed in the table. With the aid of a matching screw, which regulated the position of the coupling loop in the resonator, it was possible to vary the gain from 5 to 38 db. The bandwidth accordingly decreased in this case. The amplifier, operating at 10.8 gigacycles, had a bandwidth on the order of 300 mc at K = 5 db and 8 mc at K = 28 db.

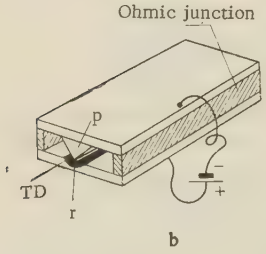
The tested diodes were made of gallium-arsenide with specific resistivity $\rho = 15 \cdot 10^{-4}$ ohm-centimeters. The value of I_{max} ranged from 0.5 to one milliampere, and the ratio I_{max}/I_{min} was ≥ 3.

In addition to the amplifier circuits considered above, great interest attaches also to distribution systems, which were theoretically examined in reference [15].

Fig. 22a shows one possible model of a distributed tunnel diode, in the form of a strip line, with which it is possible to realize a traveling wave amplifier. As shown by preliminary calculations, such an amplifier will have a sufficiently high frequency limit and a large wave resistance only at a very low width (W) of the tunnel diode (see Fig. 22).

One must therefore regard as the more promising the distributed-amplifier model, shown in Fig. 22b, where structurally the p-n junction is made in the form of a "knife-edge" contact, located in the middle of a strip line. As shown by preliminary analysis, the fundamental energy in such an amplifier will be concentrated in the central part of the line, where the p-n





junction is located. Therefore it is possible to connect the constant bias source without special precautions on any edge of the line, as shown in Fig. 22b.

Fig. 22. Model of distributed amplifier using a tunnel diode:
a — tunnel diode with flat contact;
b — tunnel diode with knife-edge contact.

In addition to the difficulties of technological character, connected with the preparation of complex p-n tunnel junctions, in the development of distributed amplifiers using tunnel diodes the serious problem arises of stability of such systems, this problem being directly connected with the conditions of matching over a wide frequency range.

The authors are deeply grateful to V.V. Migulin for many valuable hints and remarks.

REFERENCES

1. L. Esaki. New phenomenon in narrow Ge p-n junction, Phys. Rev., 1958, 109, 603.
2. C. Mead. A note on tunnel emission, Proc. IRE, 1960, 48, 8, 1478.
3. J.V. Morgan, E.O. Kane. Observation of direct tunneling in Ge, Phys. Rev. Letters, 1959, 3, 10, 446.
4. V.L. Bonch-Bruevich. On the theory of the tunnel diode. Radiotekhnika i elektronika [Radio Engineering and Electronic Physics], 1960, 5, 12, 2033.
5. E.O. Kane. Theory of tunneling, J. Appl. Phys., 1961, 31, 1, 83.
6. I.A. Lesk, K.N. Holonyak, U.S. Davidson, M.W. Aarons. Ge and Si tunnel diodes design, operation and application, Wescon Convention Record IRE, 1959, 3, 3, 9.
7. R. Trambarulo, C.A. Burrus. Esaki diode microwave oscillators, Proc. IRE, 1960, 48, 10, 1776.
8. I.C. Green, E.W. Sard. Experimental tunnel diode mixer. Proc. IRE, 1961, 41, 1, 350.
9. E. Goto, et al. Esaki diode high-speed logical circuits. IRE Trans. Electronic comput., 1960, EC-9, 1, 25.
10. F.P. Heiman. 100-mc tunnel-diode ring counter. Proc. IRE, 1961, 49, 7, 1215.
11. H.S. Sommers. Tunnel diodes as high-frequency devices. Proc. IRE, 1959, 47, 7, 1206.
12. C.A. Burrus. Millimeter wave Esaki diode oscillators. Proc. IRE, 1960, 48, 12, 2024.
13. A.A. Andronov, A.A. Vitt, S.E. Khaikin. Teoriya kolebanii [Theory of Oscillations], GIFML, 1959.
14. U.S. Davidsohn, Y.C. Hwang, G.B. Ober. Designing with tunnel diodes, parts 1, 2. Electronic design, 1960, 3, 50.
15. M.E. Hines. High-frequency negative-resistance circuit principles for Esaki diode applications. Bell System Techn. J., 1960, 39, 3, 477.
16. R.P. Murray. Biasing methods for tunnel diodes, Electronics, 1960, 33, 23, 92.
17. K.K.N. Chang. Low-noise tunnel-diode amplifier. Proc. IRE, 1959, 47, 7, 1268.
18. V.I. Kalinin, G.M. Gershtein. Vvedeniye v radiofiziku [Introduction to Radiophysics], GTTI, 1957.
19. G. Schaffner. A compact tunnel-diode amplifier for ultrahigh frequencies. Wescon Convention Record, IRE, 1960, 4, 2, 86.
20. L. Esaki. Fundamentals of Esaki tunnel diode in circuit applications. URSI XIII General Assembly, London, 1960.
21. A.G. Chunoweth, W.L. Feldman, R.A. Logan. Excess tunnel current in Si Esaki junctions. Phys. Rev., 1961, 121, 3, 684.
22. N.A. Belova, A.N. Kovalev. Eksperimental'noye issledovaniye tunnel'nogo toka v uzkikh germaniyevykh p-n-perekhodakh, [Experimental investigation of tunnel current in narrow p-n junctions], Radiotekhnika i elektronika, 1961, 6, 1, 160.

23. J.J. Tiemann. Shot noise in tunnel diode amplifiers. Proc. IRE, 1960, 48, 8, 1418.
24. T. Yajima, L. Esaki. Excess noise in narrow Ge p-n junction. J. Phys. Soc. Japan, 1958, 13, 11, 1281.
25. I.I. Ivanchik. K teorii vyrozhdennogo p-n perekhoda [On the theory of the degenerate p-n junction] Fizika tverdogo tela, 1961, 3, 1, 103.
26. Price, Radklif. The theory of tunnel diode, IBM J., 1959, 3, 4, 364.
27. B.C. Wilson. Tunneling current in Esaki diodes. Phys. Rev., 1961, 121, 4, 1070.
28. A.R. Cdawa. Tunnel-devices. Quart. Progr. Rept. Solid State Res. Inst. Technol., 1960, April, 4-7.
29. B.M. Vul, A.P. Shotov, S.P. Trishechkina. Temperaturnaya zavisimost' tunnel'nogo toka v p-n perekhodakh [Temperature dependence of tunnel current in p-n junctions], Fizika tverdogo tela, 1961, 3, 2, 667.
30. H.S. Sommers, Jr., K.K.N. Chang, H. Nelson, R. Stienhoff, R. Schnitzler. Tunnel diode for low noise amplification. Wescon Convention Record IRE, 1959, 3, 3, 3.
31. E.G. Nielsen. Noise performance of tunnel diodes. Proc. IRE, 1960, 48, 11, 1903.
32. P. Penfield, Jr., Noise performance of tunnel-diode amplifiers. Proc IRE, 1960, 48, 8, 1478.
33. La Rosa Richard, C.R. Wilhelmsen. Theoretical justification for shot-noise smoothing in the Esaki diode. Proc. IRE, 1960, 48, 11, 1903.
34. J.J. Tiemann. Tunnel diodes and their use as multifunctional circuit elements, 1960 Internat. solid-state circuits conference, N.Y. Lewis Winner, 1960, 8-9.
35. W.W. Gartner. Esaki or tunnel diode, pt 2, Semiconductor products, 1960, 3, 6, 36.
36. E. Miller, H.B. Yin, J.B. Schultz. Tunnel-diode radio-frequency amplifier, Wire and Radio Communs, 1960, 78, 10, 48.
37. H.A. Haus, R.B. Adler. Optimum noise performance of linear amplifiers. Proc. IRE, 1958, 46, 8, 1517.
38. K.K.N. Chang. The optimum noise performance of tunnel-diode amplifiers, Proc. IRE, 1960, 48, 1, 107.
39. J.J. Tiemann. Shot noise in tunnel-diode amplifiers. Proc. IRE, 1961, 49, 3, 622.
40. A.V. der Ziel. Noise of measure of lossy tunnel-diode amplifiers. Proc. IRE, 1960, 48, 7, 1321.
41. J.J. Sie. Absolutely stable hybrid coupled tunnel-diode amplifiers. Proc. IRE, 1960, 48, 7, 1321.
42. L.A. Lo Sasso. Tunnel diode as an interstage gain device, Proc. IRE, 1960, 48, 4, 793.
43. E.W. Sard. Tunnel (Esaki) diode amplifiers with unusually large bandwidths, Proc. IRE, 1960, 48, 3, 357.
44. E.D. Long, C.P. Womack. Designing tunnel diode R-F amplifiers. Electronics, 1961, 34, 7, 120.
45. A. Yariv, J.S. Cook. Operation of an Esaki diode microwave amplifier. Proc. IRE, 1960, 48, 6, 1155
46. W.J. Robertson. A broad-band hybrid doupled tunnel diode down converter. Proc. IRE, 1960, 48, 12, 2023
47. R.F. Trambarulo. Esaki diode amplifiers at 7, 11, and 26 kMc. Proc. IRE, 1960, 48, 12, 2022.
48. Tunneldiode- und Reaktanzverstarker fur das VHF- und UHF-Gebiet, Electron Rundschau, 1961, 15, 3, 121.
49. S.B. Geller, P.A. Mantek. Tunnel diode large-signal simulation study. Proc. IRE, 1961, 49, 4.
50. H.R. Kaupp, D.R. Crosby. Calculated waveforms for tunnel diode locked pair. Proc. IRE, 1961, 49, 1, 146.
51. M. Schuller, W.W. Gartner. Large-signal circuit theory of negative resistance diodes, in particular tunnel diodes. Proc. IRE, 1961, 49, 8, 1268.

Received by editor 7 June 1961

SEQUENTIAL METHODS OF CORRECTING SINGLE AND MULTIPLE CLOSELY-SPACED ERRORS IN GROUP CODES

A.N. Radchenko and E.T. Mironchikov

A method is proposed of constructing simple sequential devices for the correction of several closely-spaced errors in one class of group codes. The method consists of having the controlled code word pass through two parallel delay lines. One line, which has logical feedback, passes only the random noise pulses. The other line serves for time matching. Its output is connected to the output of the first delay line by means of a non-coincidence circuit, which restores the distorted pulses of the code word. Examples of corrector circuits for the correction of errors are given.

Natural, industrial, and other noise generated by a single source are as a rule correlated. Usually their action results in distortion of a group of closely-spaced symbols of the code work. The distortion of several symbols that are far from each other is much less probable. Such noise includes, for example, thunderstorm discharge, fading in radio reception, etc. We shall henceforth assume that multiple errors do not go beyond the limits of a comparatively small arbitrarily situated portion of the code word.

By way of coding and decoding units we employ circuits made up of shift registers and logical feedback. The first to propose the use of shift registers with linear feedback for correction of errors in codes was Huffman [1]. He pointed out the possibility of using such circuits for the correction of single and several double errors. However the possibility of correcting grouped errors and a large number of independent errors were not investigated, nor were circuit solutions proposed.

We consider below a method of correcting grouped errors in binary codes and propose a method for the synthesis of units that correct errors in binary codes. In addition, we present an analysis of such devices.

Let us consider certain properties of shift registers with logical feedback. The diagram of such a register is shown in Fig. 1. The letters D denote the cells of the register. In the simplest case they can be represented by single delay lines. The logical feedback consists of two branches. The first is a linear function which can be realized with the aid of non-correspondence circuits, denoted by the symbol $\oplus$ (modulo-2 adders). A block of such circuits serves as a parity indicator. The output circuit of the parity indicator is excited if an odd number of input circuits is excited. Huffman [2] proposed a mathematical formalism, suitable for the designation and transformation of linear circuits with shifting registers and modulo-2 adders, and determined many properties of such circuits. The second feedback branch is represented by a non-linear function feed, which is specified in Boolean form [3].

Let us consider first the action of the linear part of the circuit, for which we assume that the initial state of the register is 000. If we apply a pulse to the input of the circuit, then it will pass through the parity indicator into the shift register. At the instant of passage through the taps k, l, and m of the shift register, the pulse will again act on the parity indicator. As a result a periodic sequence of the code ring

X... 11101001110100...

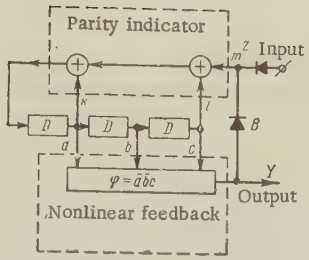


Fig. 1. Diagrams of selective delay line with logical feedback for single pulse (delay 7 steps)

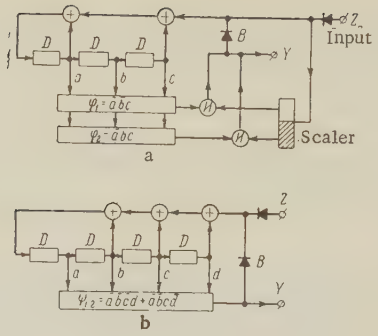


Fig. 2. Diagrams of selective delay line with logical feedback for single and doubled pulses (delay 7 setps)

will appear in the shift register. All the three-part code combinations, except 000, are represented in this code ring.

Let us consider now a nonlinear circuit φ, set for one of these combinations. Depending on the form of the combination, we can obtain an output pulse Y, which lags the input pulse by any number of steps, but by not more than the period of the ring. To obtain maximum delay it is necessary to select the last combination of the code ring. In the specific case (Fig. 1) this combination is 100, and consequently, in the language of circuit theory, $\varphi = \bar{a}\bar{b}c$. The output signal Y is applied through the diode B to the parity indicator and converts the circuit to the state 000 in place of 100. Thus, the periodic sequence X will terminate and the register will return to its stationary state (000). We note that with the aid of the circuits considered one can create economical delay lines. For example, with a 10 microsecond delay line it is possible to obtain delays up to 1023 microseconds.

It is possible to transmit with delay not only single pulses, but also pulses that follow each other (doubled). For such an expansion of the capabilities of the circuit we use the property of maximal code rings, which consists in the fact that the modulo-2 sum of two identical rings in different phases (i.e., shifts relative to each other) yields the same code ring [2].

Because of the linear properties of the shift registers, the action of the doubled pulses can be regarded as the action of two independent pulses, and the result can be considered as the sum of code rings, generated by these pulses. For the case under consideration we have

$$\begin{array}{l} Z_1 \ldots 010000000 \ldots \\ \oplus\ Z_2 \ldots 001000000 \ldots \\ \hline Z_{1,2} \ldots 011000000 \ldots \end{array} \qquad \begin{array}{l} X_1 \ldots 011101001110100 \ldots \\ \oplus\ X_2 \ldots 001110100111010 \ldots \\ \hline X_{1,2} \ldots 010011101001110 \ldots \end{array}$$

It follows therefore that for maximum delay of the first of the doubled pulses it is necessary to separate from the code ring the combination 110, i.e., $\varphi = \bar{a}bc$. The corresponding circuit is shown in Fig. 2a. When a single pulse is applied to the input, this circuit behaves like the circuit of Fig. 1. The difference lies only in that the selector φ_1 is connected to the input via an AND circuit. The latter is controlled by a binary scale, which controls the parity of the input signal. If a single pulse is applied to the input, the scale turns on the selector φ_1; if the pulse is doubled, φ_2 is connected. When doubled pulses are applied to the input, code ring $X_{1,2}$ is duplicated until the combination of the code ring 110 is selected. The signal from selector φ_2 is applied to the input Y and simultaneously will act through the diode B on the block of adders in modulo-2. As a result, instead of the next combination 101, the combination 100 will appear on the register. Simultaneously the pulse from Y will enter the scale, and this will cause Y to switch over to selector φ_1. Since the register contains at that instant the combination 100, the second delayed pulse will immediately be delivered. Thus, the circuit can pass both single and doubled pulses. Circuits of this type will be called selective delay lines (SDL), and the totality of the pulse combinations passed by such systems will be called the transmission characteristic of the SDL.

It is possible in principle to replace the scale by a shift-register cell and combine it with

the main register, thus obtaining a circuit having the same transmission characteristic. Such a circuit is shown in Fig. 2b. When a single pulse is applied to this circuit, a code ring of four-part codes is excited in it.

$$\ldots 10110001011000\ldots ,$$

and when doubled pulses are applied to the input, the circuit produces the code ring

$$\ldots 11101001110100\ldots .$$

Thus, a single pulse and doubled pulses produce in this circuit different particular code rings [3]. This relieves us of the need for auxiliary commutation of the selectors, thereby greatly simplifying the circuit. Such a circuit becomes outwardly similar (see Fig. 2b) to the circuit of Fig. 1, but unlike the latter it distinguishes a single input pulse from a doubled one and accordingly reproduces them with a delay of seven steps. This leads to a simple method of increasing the number of combinations transmitted by the selective delay line: Each combination which should be delayed by n steps, should excite in the circuit its own particular code ring of length n symbols.

To determine the cyclic properties of shift registers with feedback, we can use the methods of the theory of finite fields [4]. For this purpose, the linear part of the circuit is described by a polynomial of the algebraic delay operator D with coefficients from the field of modulo-2 residues. Thus, the linear part of the circuit shown in Fig. 2b can be described by the polynomial $F(D) = 1 \oplus D^2 \oplus D^3 \oplus D^4$. This notation indicates that the output signals from the second, third, and fourth cells of the shift register are added in modulo-2 and are written in the first cell. The polynomial F(D) can be resolved into factors that are irreducible in the field of residues in modulo 2. The cyclic properties of the irreducible factor of degree k_i are written in the form of the sum

$$\{1 + \mu_i(n_i)\}, \tag{1}$$

where n_i — length of the code ring, corresponding to the i-th factor; $\mu_i = (2^{k_i} - 1)/n_i$ — number of code rings of length n_i at the i-th factor; 1 indicates a cycle consisting of zeros only.

The cyclic properties of the shift register with feedback, described by a reducible polynomial, are determined by the product of the formal sums in the form (1), corresponding to the irreducible factor [4]. Let us illustrate in greater detail the design of selective delay lines using as an example a 15-step delay line, from which we require the transmission of the following pulse combinations: 100, 110, 101, 111. The sum sought should have four

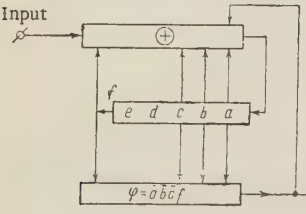


Fig. 3. Diagram of selective delay line with logical feedback for a group of arbitrarily located pulses (delay by 15 steps).

code rings, each 15 symbols in length, and to each of the foregoing combinations there should correspond its own code ring. Considering the expansion $D^{15} \oplus 1$ into factors irreducible in the field of modulo-2 residues:

$$D^{15} \oplus 1 = (D^4 \oplus D^3 \oplus 1)(D^4 \oplus D \oplus 1)(D^4 \oplus D^3 \oplus D^2 \oplus D \oplus 1) \times \times (D^2 \oplus D \oplus 1)(D \oplus 1), \tag{2}$$

we can note that the sought polynomial will consist of two factors:

$$F(D) = (D^4 \oplus D^3 \oplus 1)(D^2 \oplus D \oplus 1) = D^6 \oplus D^3 \oplus D^2 \oplus D \oplus 1.$$

The cyclic properties of each factor are determined by the sums

$$\{1 + \mu_1(n_1)\} = \{1 + 1\,(15)\},\ \{1 + \mu_2(n_2)\} = \{1 + 1\,(3)\}.$$

The product of these sums yields the cyclic properties of the selected shift register with

feedback

$$\{1 + \mu_1(n_1)\}\{1 + \mu_2(n_2)\} = \{1 + \mu_1(n_1) + \mu_2(n_2) + \mu_1\mu_2\ \mathbf{LCD}[n_1, n_2]\times$$
$$\times\ (\mathbf{SCM}\ [n_1, n_2])\} = \{1 + 1(15) + 1(3) + 3(15)\} = \{1 + 1(3) + 4(15)\},$$

where LCD $[n_1, n_2]$ is the greatest common divisor of the numbers n_1 and n_2; SCM $[n_1, n_2]$ is the lowest common multiple of the numbers n_1 and n_2. The selected shift register with feedback reproduces four 15-member code rings, and with respect to the specified pulse combinations, we have the following one-to-one correspondence:

1) 100 — 110011100100000,

2) 110 — 101010010110000,

3) 101 — 111111011101000,

4) 111 — 100110101111000.

The underlined combination must be selected by means of the circuit

$$\varphi = \bar{a}\bar{b}\bar{c}\bar{d}\bar{e}f \cup a\bar{b}\bar{c}\bar{d}\bar{e}f \cup \bar{a}\bar{b}\bar{c}\bar{d}ef \cup a\bar{b}\bar{c}\bar{d}ef = \bar{a}\bar{b}\bar{c}f.$$

A delay line circuit transmitting pulse combinations of length ≤ 3 is shown in Fig. 3. Circuits that insure delay of different pulse combinations on a larger segment are analogously constructed. It is typical that during the process of reproduction of the delayed combination of the pulses, the nonlinear feedback causes every time a series of transitions from one ring to the other, if successively returning the circuit to the initial state. These transitions, caused by the output signals, duplicate the action of the input signal, but cause a change in the code rings in a different sequence. In spite of the complexity of the description of the processes, the realizing circuits are quite simple and effective.

The examined class of circuits has a very important property. It turns out that certain combinations of pulses are not at all transmitted by such circuits. For example, with respect to the circuit of Fig. 1, such properties are possessed by the following combinations:

0000000	0110100	1101000	1011100
0001101	0111001	1100101	1010001
0011010	0101110	1110010	1001011
0010111	0100011	1111111	1001011.

The circuit of Fig. 2 does not transmit the eight combinations:

0000000	1011100
0010111	1001011
0101110	1110010
0111001	1100101.

For the circuit of Fig. 3 one can list 512 such combinations.

It is easy to verify directly that the foregoing combinations represent combinations of group codes with Hamming distance d = 3 and d = 4 for the first and second circuits respectively. Thus, each of the foregoing circuits is a special kind of blocking filter, tuned to combinations of a definite code. The number of combinations which are not transmitted by the selective delay lines is determined by the relation

$$M = \frac{2^{n-s+1}}{n+1},$$

where n is before the length of the particular code rings used to correct the errors; s is the greatest length of the transmitted pulse combination.

No less important a feature of the selective delay line is that if the operation of the nonlinear feedback is suitably operated, the circuit has superposition properties. In other words, the result of simultaneous action of two input pulse sequences (for example, a code word and a noise pulse) can be regarded as the sum of the results of two separately acting signals. Thus, if any group of pulses in the code combination is distorted by noise, then only the combination of the error pulses will appear at the output of the selective delay line.

For a code word, the selective delay line will as before be a blocking filter.

The properties of the circuits considered can be used to correct errors in group codes. Fig. 4 shows such a device. It consists of two parts, a selective delay line (SDL) and an ordinary delay line (DL). The first line, having the logical structure considered above, passes only the noise pulses. The second line delays the entire code word. The delay time in these two lines is the same and equal to the length of the code word. The outputs of both parts are connected by a noncoincidence circuit (NC) (modulo-2 adder), and this makes it possible to correct errors in code words.

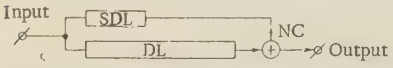


Fig. 4. Structural devices for correcting errors in a code.

The superposition property is characteristic of linear circuits only. In the case considered it is obtained artificially by suppressing the action of the nonlinear feedback element during the time of joint processing of the signal and the noise. The suppression is realized by a suppression circuit, which turns off the selective circuit during the time of code reception. By way of example, Fig. 5 shows a more detailed circuit, which corrects single errors in a 127-term code. Here the code word proceeds from the output of the receiving unit (REC) and is inserted in the memory register (MR). It is simultaneously fed to a control register (CR), which comprises a selective delay line which performs the role of a blocking filter for this code. The control register consists of seven cells of unit delay with logical feedback of two modulo-2 adders. After receiving the code word, all the cells of the control register go to the zero state, if no errors occurred.

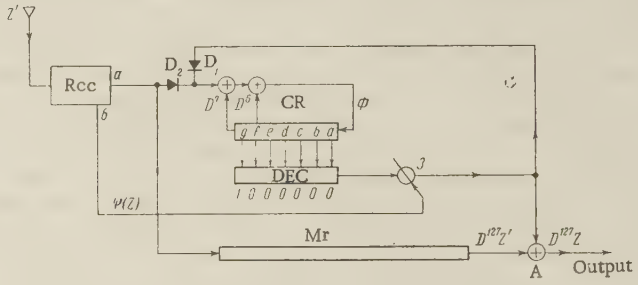


Fig. 5. Diagram of corrector for single pulses in a 127-term code.

If one of the pulses of the code word prove to be distorted, the last combination on the cells of the control register determines uniquely the location of the error. During the time of reception of the code word, a signal picked off the output b of the receiver suppresses the action of the selector DEC (decoding of the end of the cycle). Therefore during the time of the succeeding operation of the code register, the selector DEC notes the end of the cycle and issues a signal to correct the error on the adder A. This signal will be simultaneously applied through diode D_1 to the code register, which returns to the zero state. The diode D_2 prevents the correction pulse from entering into the memory register. The feedback of the selective delay line is described by the formula

$$\Phi = z \oplus f \oplus g \cup \bar{\psi}(z)\,\bar{a}\bar{b}\bar{c}\bar{d}\bar{e}fg.$$

Similarly, one constructs devices to correct closely-located or multiple errors. For this purpose it is necessary to choose a selective delay line having a transmission characteristic corresponding to the expected character of the errors. The concept used in this work is apparently simpler than the Hamming concept. This does not mean, naturally, that the differences are of principal character. It is always possible to show that the operation of a selective delay line can be reduced to Hamming tests for correspondence by a matrix of a special type.

The symbols of the code sequence can be regarded as coefficients of a polynomial of the algebraic delay operator D. Then the question of the distortion of the code word reduces to the question of the divisibility of this polynomial by the polynomial of the coding converter. The polynomial in D corresponding to the undistorted code word is completely divisible by the polynomial of the coding converter. The polynomial corresponding to a distorted code word generates as a result of such a division an infinite periodic sequence of symbols of the code

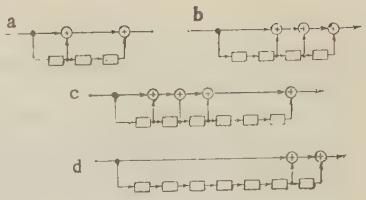


Fig. 6. Converter of information into a noise-immune code:
a — to obtain a seven-term code with correction of one error (information capacity 2^4);
b — to obtain seven-term code with correction of one or two neighboring errors (information capacity 2^3);
c — to obtain 15-term code with correction of errors of the type X, XX, XOX, XXX on an arbitrary situated section (information capacity 2^9).

ring, terminated only by the action of the nonlinear feedback. The action of the matter is equivalent to an artificial restoration of the divisibility of the polynomial by means of the formula

$$\frac{N(D) \oplus D^n(N(D))}{F(D)}, \tag{3}$$

where n is the length of the code ring; N(D) corresponds to the combination of error pulses; $D^n(N(D))$ corresponds to the combination of pulses formed in the nonlinear feedback circuit, i.e., the output combination of the selective delay line. It follows from (3) that the operation of correcting the errors is delayed by n steps. An analysis of the processes occurring in the selective delay line upon correction of the errors will be made using as an example the line shown in Fig. 3. Its connection is shown in Figs. 4 and 5. We represent the transmitted code word

$$Z = 110101011000111$$

in the form of a polynomial in D:

$$Z(D) = 1 \oplus D \oplus D^3 \oplus D^5 \oplus D^7 \oplus D^8 \oplus D^{12} \oplus D^{13} \oplus D^{14}.$$

Assume that in the received code word the symbols distorted are in the seventh and ninth positions. Then the received sequence can be written

$$Z'(D) = Z(D) \oplus N(D) = 1 \oplus D \oplus D^3 \oplus D^5 \oplus D^6 \oplus D^7 \oplus D^{12} \oplus D^{13} \oplus D^{14},$$

where N(D) corresponds to a combination of error pulses. The processes occurring in the selective delay line upon reception of a distorted code word are described by the relation

$$T(D) = \frac{Z(D) \oplus N(D) \oplus D^n(N(D))}{F(D)}.$$

For the example considered here we have

$$T(D) = \frac{1 \oplus D \oplus D^3 \oplus D^5 \oplus D^6 \oplus D^7 \oplus D^{12} \oplus D^{13} \oplus D^{14} \oplus D^{15}(D^6 \oplus D^8)}{1 \oplus D \oplus D^2 \oplus D^3 \oplus D^6} =$$
$$= 1 \oplus D^2 \oplus D^3 \oplus D^5 \oplus D^9 \oplus D^{10} \oplus D^{11} \oplus D^{13} \oplus D^{14} \oplus D^{15} \oplus D^{17}.$$

Going again to binary sequences, we obtain

$$T = 10110100011101110\underline{1000}00\underline{1}.$$

The selected combinations 101000 and 100000 correspond to two errors. Using this example we can trace the transition of the selective delay line from one code ring to the other and then to the zero position.

The code words are formed also with the aid of a shift register with logical connections. The coding circuits for all the cases cited in this paper are shown in Fig. 6.

Conclusions:

1. By connecting the intermediate taps of the delay line to its input through logical

networks, we can increase the delay time (by reducing the number of pulse combinations transmitted by the line).

2. From among the set of pulse combinations of length n we can separate two subsets: the combination of the first passes through the selective delay line without distortion; the combinations belonging to the second subset do not pass through the line at all.

3. The selective properties of the selective delay line can be used to correct single and multiple errors in group codes. For this purpose it is necessary to employ delay lines whose transmission characteristic corresponds to the expected character of the errors.

4. The group code consists of the totality of combinations which do not pass through the selective delay line. The polynomial characterizing this totality is the divisor of the polynomials corresponding to these combinations. Naturally, the effect of the coding converter reduces to a multiplication of a set of polynomials, characterizing the combinations made up of information symbols, by the polynomial of the coding converter.

5. The resultant sequential corrector networks are much simpler than the known combinational sequences.

REFERENCES

1. D.A. Huffman. IRE Trans., 1956, IT-2, September, 20.
2. D.A. Huffman. Synthesis of Linear Sequential Coding Networks. Department of electrical engineering and research lab of electronics, MIT, Cambridge, Mass., 1955.
3. A.N. Radchenko. Methods of synthesizing code rings. Radiotekhnika i elektronika, 1959, 4, 11, 1783.
4. B. Elspas, IRE Trans., 1959, CT-6, March, 45.

Received by editors 16 January 1961

TRANSFORMATION OF RANDOM SIGNALS IN NONLINEAR LINES

S.A. Akhmanov and R.V. Khokhlov

The problem of the passage of a weak signal with random amplitude and phase through a nonlinear line, excited simultaneously by a signal of finite amplitude, is investigated theoretically. The one-dimensional laws of distribution of the amplitude and phase of the weak signal in various sections of the line are investigated in detail, with the frequencies of the finite-amplitude being close to double the frequency of the weak signal, and for small nonlinearity. It is shown that the phase of the weak signal at the output can be located in only one of two rather narrow regions, the centers of which are a distance π apart. The conditions are derived under which the relative times of stay of the phase in these regions are determined exclusively by the statistical characteristics of the phase at the input of the line. The possibilities of using this circumstance for the design of a phase filter are discussed.

INTRODUCTION

Nonlinear waveguide systems, particularly systems with nonlinear reactive parameters, are acquiring an ever growing importance in radio. In the theoretical papers published to date, principal attention has been paid to problems connected with the propagation and interaction of regular signals in nonlinear waveguide systems [1, 2].

Considerable interest is attached to analogous problems pertaining to random signals, signals with random amplitude or phase. The present paper is devoted to an analysis of the passage of a weak random signal through a nonlinear transmission line, excited simultaneously by a regular signal of finite amplitude.* It is natural to expect that the action of the finite-amplitude signal can lead to an appreciable modification of the statistical characteristics of the amplitude and phase of the random signal; the latter may be of interest from the point of view of the problem of filtering signals. In the case of small nonlinearity and low losses, the equations that describe the amplitude and phase of the wave propagating through the nonlinear line are analogous to the abbreviated equations widely used in oscillation theory [3]. In the problem considered here, the equations for the amplitude and phase of the weak signal are essentially symbolic differential equations, since for the case of greatest interest, that of very weak signals, an account must be taken of the intrinsic thermal noise (fluctuation of external force), distributed along the line. In the present paper we analyze both the case when the interaction of the inherent noise of the line lead only to small fluctuations in the amplitude and phase of the propagating wave, and the case when large phase fluctuations become appreciable.

1. DERIVATION OF FUNDAMENTAL EQUATIONS

Let us consider a nonlinear transmission line, the equivalent circuit of an elementary cell which is shown in Fig. 1.

The distributed capacitance and the distributed conductance of the line are functions of voltage, so that the connection between the running charge and the voltage are described by the $Q(V) = CV + DV^2$, and $G = G\,F\,(V)$ where $F_0(V)$ will be defined in detail below. We shall note for the time being that at small voltages the function $F_0(V)$ is equal to unity ($F_0(0) = 1$).

*By finite amplitudes we mean amplitudes sufficient to cause noticeable nonlinear effects.

In parallel with the conductance G is connected a current generator I, which describes the thermal noises generated by the conductance G. The differential equations describing the current and the voltage in the line have the form

$$-\frac{\partial V}{\partial z}=L\frac{\partial J}{\partial t}, \quad -\frac{\partial J}{\partial z}=\frac{\partial Q}{\partial t}+GV+I(t, z). \tag{1}$$

The correlation function I(t, z) has the form [4]

$$\overline{I(t, z)\, I(t+\tau, z+s)}=\frac{2kT}{\pi}\delta(s)\int_0^\infty G_0(\Omega)\cos\Omega\tau\, d\Omega. \tag{2}$$

In the integrand of (2) we use G_0 in place of G, since an account of the line's own noise is of greatest importance in the case of weak signals.

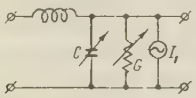


Fig. 1. Equivalent circuit of elementary cell of nonlinear line.

It is convenient to change from the equations in (1) to a simple second-order equation. Eliminating the current I, we obtain

$$\frac{\partial^2 V}{\partial z^2}-L\frac{\partial^2 Q}{\partial t^2}-L\frac{\partial GV}{\partial t}-L\frac{\partial I}{\partial t}=0. \tag{3}$$

Assume that the inputs to the line are a relatively weak signal with carrier frequency ω and a signal of frequency 2ω ("pumping") with slowly varying amplitude and phase, so that at the input (Z = 0) the voltage is given by

$$V=V_0(t)\sin[\omega t+\varphi_0(t)]+V_p\sin 2\omega t, \tag{4}$$

where $V_0(t)$ and $\varphi_0(t)$ are slowly varying functions of the time. We note that the representation of the voltage at Z = 0 in the form (4) does not exclude the possibility of analyzing cases in which the signal and pumping frequencies are not exact multiples. In these cases, the phase $\varphi_0(t)$ contains an additive term of the form $\Delta_1 t$, where Δ_1 is the deviation of signal frequency from half the pumping frequency.

Let the phase velocity in the line under consideration be weakly dependent on the frequency as the latter varies from zero to 2ω, and let it change strongly as the frequency is further increased. Under these conditions only waves with frequencies not greater than 2ω can effectively interact.

We introduce new variables

$$\zeta=t+\frac{z}{u}, \quad \eta=t-\frac{z}{u}, \tag{5}$$

where u — phase velocity at the pumping frequency. Then if the forward waves predominate in this line, and the nonlinearity coefficient D and the conductivity G are small, we can represent the solution of (3) in the form

$$V=V(\eta, \varepsilon\zeta), \tag{6}$$

where ϵ is a small parameter; $DV/C\sim\epsilon$ and $G_0/\omega C\sim\epsilon$. Changing to the coordinates ζ, η in equation (3), taking into account the form of the solution (6), and neglecting quantities of second order in ϵ, we obtain

$$-\frac{4}{u^2}\frac{\partial^2 V}{\partial\zeta\,\partial\eta}+\left(\frac{1}{u^2}-LC\right)\frac{\partial^2 V}{\partial\eta^2}-LD\frac{\partial^2 V}{\partial\eta^2}-L\frac{\partial GV}{\partial\eta}=L\dot{I}\left[\frac{u}{2}(\zeta-\eta);\frac{1}{2}(\zeta+\eta)\right]. \tag{7}$$

In the new variables, the correlation function of the random force has the form

$$\overline{\dot{I}(\zeta_1, \eta_1)\dot{I}(\zeta, \eta)} = \frac{2kT}{\pi}\delta\left[\frac{u}{2}(\zeta_1 - \zeta + \eta - \eta_1)\right] \times$$

$$\times \int_0^\infty \Omega^2 G_0(\Omega) \cos\frac{\Omega}{2}(\zeta_1 + \eta_1 - \zeta - \eta)\, d\Omega. \tag{8}$$

Subject to boundary conditions (4), the general form of the solution (6) can be made more specific

$$V = V_1(\varepsilon\zeta, \varepsilon\eta)\sin\Phi + V_p \sin 2\omega\eta, \tag{9}$$

where

$$\Phi = \omega\eta + \varphi(\varepsilon\zeta, \varepsilon\eta). \tag{10}$$

Neglecting the reaction of the signal on the pumping wave, and also its attenuation, * we assume V_p = const. Substituting (9) in (7), neglecting terms of order ϵ^2, and multiplying the equation thus obtained by cos Φ and sin Φ and averaging its terms then over the period $2\pi/\omega$, we can obtain the following abbreviated equations, which describe the behavior of the slowly-varying amplitude V_1 and phase φ:

$$\frac{\partial V_1}{\partial\zeta} - \beta V_p V_1 \cos 2\varphi + \delta(V_1)V_1 + F_1(\zeta) = 0, \tag{11a}$$

$$\frac{\partial\varphi}{\partial\zeta} + \gamma + \beta V_p \sin 2\varphi + \frac{1}{V_1}F_2(\zeta) = 0. \tag{11b}$$

Here

$$\beta = \frac{LDu^2\omega}{4}; \quad \gamma = \frac{\omega}{4}u^2\left(\frac{1}{u^2} - LC\right);$$

$$\delta(V_1) = \frac{G_0 L u^2}{4\pi}\int_0^{2\pi} F_0(V_1 \sin\Phi)\sin^2\Phi\, d\Phi. \tag{12}$$

The parameter γ enables us to take into account the weak dispersion in the line in range $0 - 2\omega$. The random forces $F_1(\zeta)$ and $F_2(\zeta)$ are

$$F_{1,2} = \frac{Lu^2}{2\omega}\frac{\omega}{\pi}\int_{-\frac{\pi}{\omega}}^{\frac{\pi}{\omega}} \dot{I}\begin{Bmatrix}\cos(\omega\eta + \varphi)\\ \sin(\omega\eta + \varphi)\end{Bmatrix} d\eta. \tag{13}$$

It is convenient to represent $F_{1,2}$ in the form

$$F_{1,2} = \bar{F}_{1,2} + f_{1,2}(\zeta); \quad \overline{f_{1,2}(\zeta)} = 0. \tag{14}$$

The mean values of the random forces, and also the statistic characteristics of the functions $f_{1,2}(\zeta)$, can be determined by using the examples developed in the theory of fluctuations of self-oscillating systems [5, 6].

We have

$$\bar{F}_1 = \frac{kT}{2\pi^2}\frac{G_0}{C^2}\frac{\omega}{u}\frac{1}{V_1}, \quad \bar{F}_2 = 0,$$

$$\overline{f_1(\zeta)f_2(\zeta_1)} = 0, \quad \overline{f_1(\zeta)f_1(\zeta_1)} = \overline{f_2(\zeta)f_2(\zeta_1)} = \frac{kT}{2\pi^2}\frac{G_0}{C^2}\frac{\omega}{u}\delta(\zeta - \zeta_1). \tag{15}$$

When representing the correlation functions in the right halves of the last equations in the form of δ-functions, we took it into account that the constant $1/\beta V_p$, which characterizes the inertia of the phase relative to the variable ζ, is much greater than $1/2\omega$, which characterizes the correlation interval of the functions $f_{1,2}$.

---

*The latter is possible if the dispersion of the conductance G is suitably chosen.

Equations (11) must be solved under boundary conditions corresponding to (4):

$$V = V_0(\eta), \quad \varphi = \varphi_0(\eta) \text{ when } \zeta = \eta. \tag{16}$$

Equations (11) are nonlinear symbolic equations. In general they are quite difficult to solve. We therefore consider first the case of a "noiseless" line, when the external fluctuation forces $F_{1,2}$ are equal to zero.

2. CHARACTERISTICS OF THE AMPLITUDE AND PHASE OF A WEAK SIGNAL IN A NOISELESS LINE

In the absence of external forces, the equation (11b) for the phase can be solved independently of the equation for the amplitude; to the contrary, the character of the variation of the amplitude depends appreciably on the behavior of the phase. The character of the solution of the phase equation is determined by the ratio of the parameters βV_p and γ [7]. When $|\gamma| > \beta V_p$ the phase is a periodic function of ζ — this corresponds to space beats in the line. When $\beta V_p > |\gamma|$ the phase tends with increasing ζ in aperiodic fashion to a value φ_{ss}, defined by the relations

$$\sin 2\varphi_{ss} = -\frac{\gamma}{\beta V_p}, \tag{17}$$

$$\cos 2\varphi_{ss} > 0. \tag{18}$$

The last case is of greatest interest, since the propagation of the signal along the line can be accompanied in this case by an exponential build up in the amplitude. We shall therefore confine ourselves to this case in the future. It follows from (17) and (18) that in the interval $-\pi < \varphi \leq \pi$ there exist two stable phases which shifted by π relative to each other. The choice of the particular stable state is determined, in the absence of external forces, exclusively by the boundary conditions. The regions of attraction of the stable states on the lines of the boundary phases are separated by the roots of the equation (17), in which $2\varphi < 0$. We shall call these regions, for brevity, regions I and II.

The solution of the phase equation (11b) is particularly simple when the phase velocities of the pumping and signal waves are equal ($\gamma = 0$). For this case we have

$$\varphi = \operatorname{arc\,tg}\left[e^{2\beta V_p(\eta-\zeta)} \operatorname{tg} \varphi_0(\eta)\right] = \operatorname{arc\,tg}\left[e^{-\beta V_p \frac{z}{u}} \operatorname{tg} \varphi_0\left(t - \frac{z}{u}\right)\right]. \tag{19}$$

If the condition

$$\beta V_p \cos 2\varphi_{ss} > \delta(0), \tag{20}$$

is satisfied, then at distances from the input

$$z_0 = \frac{u}{\sqrt{(\beta V_p)^2 - \gamma^2}}, \tag{21}$$

where the phase reaches a value close to the steady state, the amplitude begins to grow in accordance with the equation

$$\frac{\partial V_1}{\partial \zeta} - \beta V_p V_1 \cos 2\varphi + \delta(0) V_1 = 0. \tag{22}$$

Solution of this equation has in the case $\gamma = 0$ the form

$$V_1 = V_0\left(t - \frac{z}{u}\right)\sqrt{\cos^2\varphi_0\left(t - \frac{z}{u}\right) + \sin^2\varphi_0\left(t - \frac{z}{u}\right)e^{-8\beta V_p \frac{z}{u}}} \times \\ \times \exp\left[\beta V_p - \delta(0)\right]\frac{2z}{u}. \tag{23}$$

In real conditions, naturally, an unlimited growth of the amplitude, as predicted by (23),

does not take place. The factors that limit the exponential growth of the amplitude may be either the reaction of the signal wave on the pumping wave (this case is considered in [3]), or the change in the phase velocity of the growing wave with increasing amplitude, as a result of which the synchronism between the phase velocities of the signal and pumping waves* is violated, or finally nonlinear absorption. Let us consider this last case. For simplicity we can assume, for example,

$$\delta(V_1) = \delta(0)\left[1 + \frac{V_1^2}{V_p^2}\right], \tag{24}$$

where V_p is a parameter that characterizes the level of "operation" of nonlinearity. By the same token we make the dependence of the conductance on the voltage $G = G_0F(V)$ specific.

The steady-state amplitude V_{ss} is determined from the relation

$$\beta V_p \cos 2\varphi_{ss} = \delta(0)\left[1 + \frac{\Gamma_{уст}^2}{\Gamma_p^2}\right]. \tag{25}$$

The results obtained enable us to analyze the passage of a signal with random-modulated amplitude and phase through our system. It is important to emphasize, first of all, that whereas the amplitude and phase of the random signal are random functions of a single variable — the time t — at the input, they are random functions of two variables, namely t and z, in the line, and the statistical characteristics relative to t may be quite different from those relative to z. We shall assume that $V_0(t)$ and $\varphi_0(t)$ are stationary random processes and that at $z = 0$ their one-dimensional distribution laws and correlation functions are specified. It is seen from (19) and (23) that the amplitude V_1 and the phase φ are inhomogeneous random functions of z and stationary functions of t. The one-dimensional distribution function of the phase $W_z(\varphi)$ can be readily calculated by the rule for the change of variables, if we know $W_0(\varphi_0)$. Thus, for example, when $\gamma = 0$ and $W_0(\varphi_0) = 1/2\pi$ we have

$$W_z(\varphi) = \frac{1}{2\pi}\frac{e^{\beta V_p \frac{z}{u}}}{\cos^2\varphi + e^{2\beta V_p \frac{z}{u}}\sin\varphi}. \tag{26}$$

It is easy to see that as z increases the function $W_z(\varphi)$ becomes appreciably modified (see Fig. 2), and turns in the limit into two δ-like surges when $\varphi = 0$ and $\varphi = \pi$. The probabilities of establishing one of the two possible phases φ_{ss} will be determined exclusively by the probabilities of the initial phase φ_0 entering into the attraction regions from I and II.

The amplitude V_1, like the phase, is a homogeneous random function of z; using (19) and (22) and specifying the specific form of $W_0(V_0)$, we can calculate the statistical characteristics of V_1 in any cross section z. It is easy to see that independently of the form of $W_0(V_0)$, the one-dimensional distribution function of the amplitude V_1 tends as $z \to \infty$ to a δ-function near the steady-state value, as determined from (25). Fig. 3 shows by way of an example plots of the distribution functions $W_z(V_1/V_{ss})$ for different values of the parameter $r = 4[\beta V_p \cos 2\varphi_{ss} - \delta(0)]\frac{z}{u}$; in this case the (we assume a Rayleigh) one-dimensional distribution function for the amplitude at the input ($z = 0$) (see also [8]).

Thus, in spite of the fact that the boundary amplitude and phase run through a continuous series of values, the phase of the output signal can assume only one of two discrete values, and the amplitude at the output, with suitable choice of nonlinearity, is in practice generally independent of the input amplitude. The foregoing means that in a nonlinear waveguide system it is possible to create conditions under which the fluctuations of the input signal manifest themselves at the output only in a change in the probabilities of the relative duration of the output signal in one of two stable states, characterized by regular amplitudes and phases, and differing only in the values of the steady-state phase, i.e., simultaneously the system "processes" the signal along with amplifying it.

---

*For a theoretical analysis of this limitation mechanism it is necessary to include higher order terms (at least third order) in the expression for Q(V).

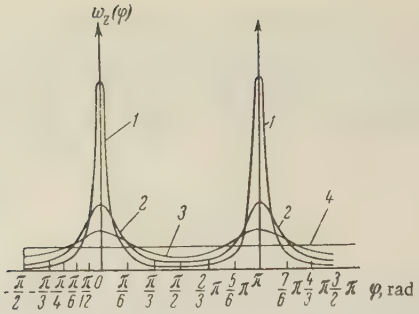


Fig. 2. One-dimensional distribution functions of the phase for different $l = 4\beta V_p z/u$ for the case when the one-dimensional distribution law of the boundary phase is uniform.
1 — $l = 2$; 2 — $l = 1$; 3 — $l = 0.5$; 4 — $l = 0$

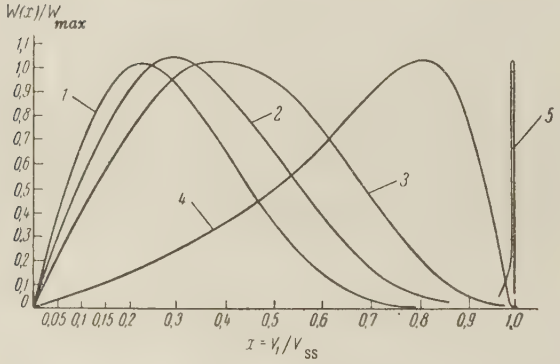


Fig. 3. One-dimensional distribution functions of the amplitude for different values of r. The ratio $V_{ss}^2/\sigma^2 = 5$; a Rayleigh one-dimensional distribution is assumed for the boundary amplitude (the only input is noise with distribution σ^2):
1 — r = 0.01; 2 — r = 0.5; 3 — r = 1; 4 — r = 2.5; 5 — r = 10

By measuring the relative durations of the output signal in the indicated stable states, we can obtain information on the distribution function of the output-signal phase, and therefore the waveguide system considered here can be used to construct a phase filter, in which a criterion for the presence of the signal is the change of the relative duration of the output phase in one of the possible states. For such a signal to operate effectively it is essential that the output state be determined exclusively by the boundary conditions; however, the internal noise of the line, generally speaking, also affects the statistics of the output states and leads in this manner to losses of information transmitted from the input. In the next section we shall consider this question in greater detail.

3. STATISTICAL CHARACTERISTICS OF THE PHASE OF A WEAK SIGNAL IN A NOISY LINE

From the point of view of the problem stated at the end of Section 2, the greatest interest is attached to an investigation of the abbreviated equation for the phase (11) with a random force. Actually, the fluctuations of the amplitude do not influence in principle the result obtained when the relative time of stay of the output signal in one of the stable states changes. The situation is different with the phase fluctuations. As can be seen from the structure of the phase equation (which is analogous essentially to the equation of motion of the Brownian particle in a potential field), the action of a fluctuation force can lead to the following:

1) "smearing" of the phase trajectories, and particularly fluctuations of the phase near the steady-state value; 2) transition of the phases from region I to II and vice versa. Obviously, small fluctuations of the phase near dynamic phase trajectories and near steady-state values do not change in principle the idealized picture described in Section 2, inasmuch as the averaged phase trajectories are determined by the boundary conditions, and the fluctuating force leads only to a smearing of these trajectories. Account of the smearing can be readily made by breaking up the phase into a regular and fluctuating parts, linearizing the equation relative to the fluctuations, and using the formalism of correlation theory. On the other hand, transitions of the phase from one region of attraction to the other play a principal role, for in this case the probability of finding the output phase in stable states will be determined not only by the boundary conditions but also by the intensity of the fluctuation force. The latter means that the processing of the signal in a noisy line is accompanied, generally speaking, by loss of information concerning the distribution function of the boundary phase. As a quantitative measure characterizing the loss of information, it is convenient to use the quantity

$$M = \frac{P_1(z) - P_2(z)}{P_1(0) - P_2(0)} \leqslant 1, \tag{27}$$

where

$$P_{1,2}(z) = \int_{I,II} W_z(\varphi)\, d\varphi; \quad P_{1,2}(0) = \int_{I,II} W_0(\varphi_0) d\varphi_0. \tag{28}$$

In a noiseless line M = 1. In a noisy line the parameter M is a function of the coordinate z, the ratio of the square-mean amplitude at the input $\sqrt{\overline{V_0^2}}$ to the intensity of the fluctuation force, etc. In the limiting case of a sufficiently intense random force, even at not very large z, we have $P_1(z) \simeq P_2(z)$ and $M \simeq 0$, since, as would follow from most general considerations, the probabilities of finding the phase in stable states have a tendency to become equalized. The quantity M can be represented in the form

$$M = \frac{\Delta P(z)}{\Delta P(0)} = \frac{\Delta P(0) - N}{\Delta P(0)} = 1 - \frac{N}{\Delta P(0)},$$

where N characterizes the equalization of the probabilities under the influence of the random force: $N \leq \Delta P(0)$.

The entire process of propagation of the signal in this line can be broken up into two stages. In the first stage the phase trajectories contract to stable states. The amplitude over the extent of this stage is practically unchanged, remaining equal in order of magnitude to the amplitude at the input of the system. The second stage is characterized by an exponential growth in the amplitude of the oscillations and an establishment of its stationary value.

It follows from (11b) that the intensity of the random action on the phase, for a specified noise level in the line, is the greater, the smaller the amplitude of the signal V_1. This means that the probabilities of the transitions are maximal on the first stage, when the amplitude is approximately V_0.

Let us confine ourselves to an estimate of the minimum amplitude of the input signal, for which the transitions are still sufficiently rare, so that the relative times of stay of the output phase in stationary states are determined essentially by the boundary conditions ($M \simeq 1$). This is precisely the case of interest if the utmost capabilities of a waveguide phase filter are to be determined. To carry out the estimate we replace the amplitude V_1 in the phase equation (11b) by its mean-squared value at the input V_0. Although such a substitution is valid to some degree only during the first stage of the establishment of the steady state, it is advantageous to examine the solution of the phase equation with constant value of V_1 for a larger interval of the variable ζ.

Thus, let us consider the equation

$$\frac{\partial \varphi}{\partial \zeta} + \gamma + \beta V_p \sin 2\varphi + \frac{1}{V_0} F_2(\zeta) = 0. \tag{29}$$

This equation is symbolic, and describes the behavior of the random phase as a function of ζ

and of the parameters. Since the external force F_2 is δ-correlated, it is advantageous to change from (29) to an equation of the Einstein-Fokker type. This has the form

$$\frac{\partial W}{\partial \zeta} + \frac{\partial S}{\partial \varphi} = 0, \tag{30}$$

where S is the probability flux through the cross section

$$S(\varphi) = -(\gamma + \beta V_p \sin 2\varphi) W - \frac{kT}{4\pi} \frac{G_0}{C^2} \frac{\omega}{u} \frac{1}{V_0^2} \frac{\partial W}{\partial \varphi}. \tag{31}$$

It is convenient to bear in mind the analogy between the establishment of a stationary distribution of the probability of the phase and the motion of a Brownian particle in a potential field U, in the form

$$U(\varphi) = \gamma\varphi - \frac{1}{2} \beta V_p \cos 2\varphi. \tag{32}$$

At the first stage there is established a quasi-stationary distribution of the phase probabilities with two sharp maxima. The length of this stage can be estimated by disregarding the action of the fluctuation forces [see Formula (21)]. The most intense transitions in the first stage occur on the section adjacent to the input, where the probability of finding the phase in the vicinity of the boundary (separatrix) between regions I and II is still relatively large. In this portion, a small random jolt towards the boundary is sufficient to overcome the "potential barrier."

At the same time, it is necessary to consider also the transitions of the phase from region I to II and back, when the one-dimensional phase distribution law already has pronounced maxima near the stable phases; the results of such an analysis enable us, in particular, to estimate the requirement imposed on the length of the system. We note that problems of the latter type (such transitions, unlike transitions of the former type, can be called "depth" transitions) have already been considered in the literature [6, 10, 11]. For specific estimates of the intensities of the transitions it is necessary to specify the expression for the distribution function of the input phase $W_0(\varphi_0)$. For the case when in addition to noise there exists at the input a weak signal

$$V_s = A \cos(\omega t + \varphi_1), \tag{33}$$

the one-dimensional distribution of the phase of the voltage, is a superposition of a signal and noise (with dispersion σ^2), has at low signal to noise ratios the form [9]

$$W_0(\varphi_0) = \frac{1}{2\pi} + \frac{A}{\sigma} \frac{1}{2\sqrt{2\pi}} \cos(\varphi_0 - \varphi_1). \tag{34}$$

Let us estimate the intensity of the phase transitions. For simplicity we confine ourselves to the case $\gamma = 0$.

At small random-force intensities, the transitions through the boundary of the attraction regions is from states close to this boundary: $\varphi \simeq \varphi_b = \pm \pi/2$. Therefore in the vicinity of the boundary we can put in Eq. (30) $2\varphi \simeq -2\varphi_2$, where $\varphi_2 = \varphi - \varphi_b$, so that positive φ_2 correspond to one region of attraction and negative ones to the other. Equation (29) thus assumes the form

$$\frac{\partial \varphi_2}{\partial \zeta} - 2\beta V_p \varphi_2 + \frac{1}{V_0} F_2(\zeta) = 0. \tag{35}$$

Solving this equation subject to boundary conditions

$$\varphi = \varphi_0 \text{ for } \zeta = \eta, \tag{36}$$

we can obtain

$$\varphi_2 = (\varphi_0 + \varphi_r) \exp\left(4\beta V_p \frac{\zeta}{u}\right), \tag{37}$$

where

$$\varphi_r = \frac{1}{V_0} \int_0^{2\frac{z}{u}} e^{-2\beta V_p \xi} F_2(\xi)\, d\xi. \tag{38}$$

The term φ_r in (37) is due to the fluctuation forces and describes the perturbations of the phase trajectories passing near the boundary. From (38) we can readily obtain the main statistical characteristics of φ_r:

$$\overline{\varphi}_r = 0,\ \overline{\varphi_r^2} = \frac{\overline{F_2^2}}{V_0^2} \frac{1}{4\beta V_p} \left(1 - e^{-8\beta V_p \frac{z}{u}}\right). \tag{39}$$

Approximating the distribution of φ_r by means of a normal distribution, we obtain

$$W(\varphi_r, z) = \frac{1}{\sqrt{2\pi \overline{\varphi_r^2}(z)}} \exp\left(-\frac{\varphi_{c.r}^2}{2\overline{\varphi_r^2}(z)}\right). \tag{40}$$

The probability of a phase with a value φ_0 at the input entering into a "foreign" region of attraction is

$$\int_{\varphi_0}^{\infty} W(\varphi_r, z)\, d\varphi_r. \tag{41}$$

At sufficiently large values of z, when $z \gtrsim \frac{1}{8} \frac{u}{\beta V_p}$ this probability is

$$\frac{1}{\sqrt{2\pi}} \int_{\frac{\varphi_0}{B}}^{\infty} e^{-y^2}\, dy, \tag{42}$$

where

$$B = \frac{\sqrt{\overline{F_2^2}}}{2V_0} \frac{1}{\sqrt{\beta V_p}}. \tag{43}$$

Let us consider the case when the phase of the input signal is optimal, so that φ_1 in (33) is equal to zero. In this case the distribution of the input phase φ_0 in the vicinity of the boundary can be represented in the form

$$W_0(\varphi_0) = \frac{1}{2\pi} + \frac{A}{\sigma} \frac{1}{2\sqrt{2\pi}} \varphi_0. \tag{44}$$

Then

$$\frac{N}{\Delta P(0)} = 2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \varphi d\varphi \int_{\frac{\varphi}{B}}^{\infty} e^{-y^2}\, dy \simeq B^2. \tag{45}$$

A discussion of this relation is given below.

The law of equalization of the probabilities of finding a phase in regions I and II for the case when the quasi-stationary distributions with two sharp maxima near the roots of equations (17) and (18) have already been established, can be determined in the following fashion.

For a stationary distribution of the phase probability, the probability flux S_+ through the potential barrier from region I to region II and the flux S_- in the opposite direction, are equal to each other (dynamic equilibrium). The values of these fluxes in the case of small intensities of the random forces and if $|\gamma| \ll \beta V_p$ are determined by the following relation (see [6])

$$S_{\pm} = \frac{1}{2\pi} \beta V_p \exp\left\{-\frac{1}{2\pi B^2}\right\}, \tag{46}$$

where B is determined by formula (43). If the probabilities of finding the representative point in regions I and II, namely $P_{1,2}$ are not equal to each other, the process of their equalization is described by the equations

$$\begin{aligned} \frac{dP_1}{d\zeta} &= \frac{1}{2} S_- (P_2 - P_1) + \frac{1}{2} S_+ (P_2 - P_1), \\ \frac{dP_2}{d\zeta} &= \frac{1}{2} S_- (P_1 - P_2) + \frac{1}{2} S_+ (P_1 - P_2). \end{aligned} \tag{47}$$

Equations (47) should be solved under the boundary conditions specified at the end of the first stage of phase establishment. Substituting in (47) the expressions for the fluxes (46) and recognizing that under the boundary conditions specified for fixed z the motion along ζ is equivalent to motion along the coordinate z, we have

$$\frac{dP_1}{dz} = \frac{\beta V_p}{2\pi u} e^{-\frac{1}{2\pi B^2}} [P_1 - P_2], \quad \frac{dP_2}{dz} = \frac{\beta V_p}{2\pi u} e^{-\frac{1}{2\pi B^2}} [P_2 - P_1]. \tag{48}$$

The length on which the equalization of the probabilities takes place is

$$L = \frac{2\pi u}{\beta V_p} e^{\frac{1}{2\pi B^2}}. \tag{49}$$

Thus, the length on which the depth transitions lead to equalization of the probabilties is considerably greater than the length on which the phase trajectories contract to the stable state when the condition

$$e^{-\frac{1}{2\pi B^2}} \ll 1, \tag{50}$$

is satisfied, i.e., when condition (50) is satisifed in a system of finite lengths, then the transitions from the vicinity of one stable state through the potential barrier into the vicinity of the second stable state are quite rare. From (45) and (50) it is seen that the conditions for the smallest frequency of the transitions from the states close to the separatrix on the phase plane, and from states situated at the minima of the potential function (33), are expressed in terms of one and the same parameter B. At not too small B^2, the roles of the transitions from states close to the separatrix and from stable states are comparable. The condition

$$B^2 \ll 1 \tag{51}$$

enables us to estimate the requirements that must be imposed on the parameters of the wave guide system and on the amplitude of the weak signal to be able to produce a phase filter. Indeed, if condition (51) is fulfilled, it is possible to realize a nonlinear wave guide system of finite length, in which the statistics of the output phase are determined only by the boundary conditions. Condition (51) becomes particularly clear if we express it in the form

$$\frac{1}{2\pi B^2} = \frac{\overline{V_0^2}}{\overline{V_{\text{therm}}^2}} \frac{\beta V_p}{\delta(0)} \gg 1, \tag{52}$$

where

$$\overline{V_{\text{therm}}^2} = \frac{2}{\pi} \sqrt{\frac{L}{C}} kT\omega. \tag{53}$$

It can be readily shown that $\overline{V^2}_{therm}$ is the mean square of the amplitude of the forward wave in the linear line, excited by its own noise, contained in a frequency band width $\sim 2\omega$. Thus, if the amplification condition $\beta V_p > \delta(0)$ is satisfied with a margin of three or four times, the statistics of the boundary conditions play a decisive role even when $\overline{V^2}_0 \simeq \overline{V^2}_{therm}$.

CONCLUSION

The results obtained show that the statistical characteristics of the amplitude and phase of a weak signal in an arbitrary cross section of a nonlinear line are determined in the general case by the statistics of the boundary conditions and by the action of the internal noise of the line, which accumulates with increasing distance. When condition (52) is satisfied, a decisive role is played by the statistics of the boundary conditions; in this case the relative times of stay of the output phase in two possible states are determined by the form of the distribution function of the input phase. Therefore a device that contains a nonlinear transmission line and a phase detector can continuously realize statistical sampling of the homogeneous distribution function of the phase of the input signal.

It must be noted that with the aid of equations (11) it is possible to analyze also the case when the signal frequency is not an exact multiple of the pumping frequency. If in this case condition (51) is satisfied, the output phase is determined here also exclusively by the boundary conditions and contains in addition to random transitions from the state I to state II and vice-versa, connected with the fluctuations of the input phase, also a periodic component with frequency Δ_1, equal to the deviation of the mean frequency of the weak signal from half the pumping frequency.

The authors are sincerely grateful to S.M. Rytov for valuable advice made during an evaluation of the paper. The authors are also grateful to R.L. Stratonovich for useful discussion of the material of Section 3.

REFERENCES

1. P.K. Tien. J. Appl. Phys., 1958, 29, 9, 1347.
2. A.P. Gaponov, G.I. Freidman. Izv. vuzov MVO SSSR (Radiofizika), 1960, 3, 1, 79.
3. R.V. Khokhlov, Radiotekhnika i elektronika, 1961, 6, 7, 1116.
4. S.M. Rytov. Teoriya elektricheskikh fluktuatsiy i teplovogo izlucheniya [Theory of electric fluctuations and thermal radiation], Academy of Sciences USSR, 1953, page 168.
5. S.M. Rytov, ZhETF, 1955, 29, 3, 304.
6. R.L. Stratonovich. Radiotekhnika i elektronika, 1958, 3, 4, 497.
7. R.V. Khokhlov. Dokl. AN SSSR, 1954; 97, 3, 411.
8. S.A. Akhmanov. Izv vuzov MVO SSSR (Radiofizika), 1960, 3, 1, 110.
9. B.R. Levin. Theory of random processes as applied to radio. Izd. Sovetskoye Radio, 1960.
10. S. Chandrasekhar. Stochastic problems in physics and astronomy (Russian translation), IL, 1947, page 117.
11. R.L. Stratonovich, P.S. Landa. Izv. vuzov MVO SSSR (Radiofizika). 1959, 2, 1, 37.

Physics Faculty,
M.V. Lomonosov
Msocow State University

Received by editor 31 March 1961

SYNTHESIS OF INHOMOGENEOUS LINES ON THE BASIS OF AN INPUT IMPEDANCE SPECIFIED IN THE FORM OF A RATIONAL FRACTION OF THE FREQUENCY

O. N. Litvinenko

A method is developed for determining the law of variation of the wave resistance of an inhomogeneous line from a specified input resistance. The use of the method is illustrated with an example of a synthesis of an inhomogeneous line filter and a shaping line.

INTRODUCTION

By synthesis of an inhomogeneous line we shall mean the determination of the variation of the wave impedance of the line from a specified input resistance. In the general case (for arbitrary input impedance) this is an exceedingly complicated problem and has not yet been solved. In the present article we consider the problem of synthesizing inhomogeneous lines for input impedances in the form of a rational fraction functions of the frequency. Such a choice of input impedance is dictated essentially by two considerations: 1) the possibility of approximating the specified function by means of a rational-fraction with any degree of accuracy within a limited frequency band; 2) the possibility of obtaining solutions in closed form.

It must also be noted that in some cases it is desirable to specify the input impedance in the form of a rational fraction, for example, in the synthesis of inhomogeneous line filters.

The question of the synthesis of inhomogeneous lines on the basis of the input impedance has been considered in the interesting paper by A. L. Fel'dshtein [1]. The results obtained by this author are valid for small values of the reflection coefficient, and essentially are qualitative in character. In the present article we develop a new method which is suitable for technical design.

1. METHOD OF SYNTHESIZING INHOMOGENEOUS LINES FOR A GIVEN INPUT IMPEDANCE

It is known that the processes in inhomogeneous lines are described by a system of equations

$$\frac{dU}{dx} = pL(x)I, \qquad \frac{dI}{dx} = pC(x)U.$$

If we change from the variable x to a new variable τ defined by

$$\tau = \int_0^x \sqrt{L(y)C(y)}\,dy$$

and denote the wave resistance of the line as a function of τ by $W(\tau)$, we obtain

$$-U' = pWI, \quad -I' = p\frac{U}{W}$$

(the prime denotes differentiation with respect to τ). It is therefore easy to find the differential equation for the input impedance of the line $Z = U/I$:

$$Z' - p\frac{Z^2}{W} + pW = 0. \tag{1}$$

Were we to know the dependence of the input impedance Z on τ, the determination of the wave impedance from (1) would entail no difficulty whatever. The problem, however, consists of finding the law of variation of the wave impedance $W(\tau)$ to fit a specified input impedance at a single point, say at the beginning of the line (when $\tau = 0$). The method which we shall use to solve this problem consists of the following.

We denoted by $Z_{k-1}(\tau)$ and $W_{k-1}(\tau)$ respectively the input and wave impedances of a certain original line. We assume that we know Z_{k-1} and W_{k-1} for any point of this line.

By subjecting Z_{k-1} to a linear-fraction transformation of the form

$$Z_k(\tau) = \rho_k(\tau)\frac{\frac{Z_{k-1}(\tau)}{\rho_k(\tau)} + \frac{p}{a_k}}{1 + \frac{Z_{k-1}(\tau)}{\rho_k(\tau)}\frac{p}{a_k}}, \tag{2}$$

we obtain the input impedance of a new line. In this expression a_k is a constant, ρ_k is a function of τ, and both a_k and ρ_k are independent of p.

It is easy to see that if Z_{k-1} is a rational-fraction function of p, then Z_k is a rational-fraction function of p, and the degrees of the polynomials of the numerator and denominator will be greater by one in this case.

Expression (2) enables us to classify the inhomogeneous lines by means of polynomials contained in the numerator and denominator of the expression for the input impedance.

We shall include in the zero class the infinitely long homogeneous line. For the sake of simplicity we shall assume its wave impedance to be $W_0 = 1$, and the input impedance is also $Z_0 = 1$ (and is independent of τ).

Putting $k = 1$ in (2), we obtain the input impedance of the line of the first class

$$Z_1 = \frac{1 + \frac{p}{a_1\rho_1}}{1 + \frac{p}{a_1}\frac{1}{\rho_1}}, \tag{3}$$

and when $k = 2$

$$Z_2 = \frac{1 + p\left(\frac{\rho_1}{a_1} + \frac{\rho_2}{a_2}\right) + \frac{p^2}{a_1a_2}\frac{\rho_2}{\rho_1}}{1 + p\left(\frac{1}{a_1\rho_1} + \frac{1}{a_2\rho_2}\right) + \frac{p^2}{a_1a_2}\frac{\rho_1}{\rho_2}} \tag{4}$$

which is the input impedance of the line of the second class, etc.

Since on going from Z_{k-1} to Z_k we increase the number of independent parameters by two (ρ_k and a_k), we can always choose these parameters at $\tau = 0$, in such a way as to obtain the required input impedance. Consequently, a_k and $\rho_k(0)$ can be regarded as known quantities.

To find the wave impedances of lines of different classes we substitute (2) in (1) and equate to zero the coefficients of like powers of p. As a result we obtain

$$W_k = \frac{\rho_k^2}{W_{k-1}}, \tag{5}$$

$$\rho_k' + a_k W_{k-1} - a_k W_{k-1} = 0. \tag{6}$$

It follows therefore that the wave impedance of a line of class n can be determined by solving a system of 2n equations of form (5) and (6) with $k = 1, 2, \ldots, n$.

From (5) and (6) we can eliminate $\rho_k(\tau)$ and obtain one equation relating $W_k(\tau)$ with $W_{k-1}(\tau)$. In this case the variation of the wave impedance is determined by solving a system of n first-order equations (k = 1, 2, ..., n).

It is convenient to consider $A_k(\tau) = \sqrt{W_k(\tau)}$ in place of $W_k(\tau)$.

For $A_k(\tau)$ we have the following equations:

$$A_k' A_{k-1} + A_k A_{k-1}' + a_k A_k^2 - a_k A_{k-1}^2 = 0,$$
$$A_k(0) = \frac{\rho_k(0)}{A_{k-1}(0)}, \quad k = 1, 2, \ldots, n; \quad A_0 = 1. \tag{7}$$

These equations solve our problem.

Thus, in order to find the wave impedance of a line of class n, namely $W_n(\tau)$ for a specified input impedance at one point, $Z_n(0)$, we must first calculate a_k and $\rho_k(0)$, and then determine $A_n(\tau)$ from (7). The wave impedance is determined as

$$W_n(\tau) = A_n^2(\tau).$$

The numbers a and $\rho_k(0)$ can be obtained by comparing the coefficients of like powers of p in the specified input impedance and in the line impedance $Z_n(0)$, expressed with the aid of formula (2) in terms of a_k and $\rho_k(0)$. We recommend, however, the use of a different method, which is more effective, particularly at large n.

Without giving the intermediate derivations, we cite the final results.

If we regard $Z_n(0)$ as a function of p, it turns out that a_k (k = 1, 2, ..., n) are equal to those values of p, at which the even part of $Z_n(0)$ vanishes. Knowing a_k, we obtain $\rho_k(0)$ from (2):

$$\rho_k(0) = Z_k(0)\,|_{p=a_k}. \tag{8}$$

Example. Find the variation of the wave impedance of an inhomogeneous line whose input impedance at $\tau = 0$ is

$$Z_2(p) = \frac{1 + 10p + 4p^2}{1 + 0.625\,p + 0.25\,p^2}.$$

1. Separating the even part of $Z_2(p)$ we equate it to zero

$$\frac{1 - 2p^2 + p^4}{(1 + 0{,}25\,p^2)^2 - (0{,}625\,p)^2} = 0.$$

From this we obtain $p_{1,2} = 1 = a_{1,2}$.

2. From formula (8) we determine

$$\rho_2 = Z_2(a_2) = 8.$$

3. From formula (2) with k = 2 we determine

$$Z_1 = \rho_2 \cdot \frac{\frac{Z_2}{\rho_2} - \frac{p}{a_2}}{1 - \frac{Z_2}{\rho_2}\,\frac{p}{a_2}}$$

or

$$Z_1 = 8\,\frac{1 + 2p - p^2 - 2p^3}{8 + 4p - 8p^2 - 4p^3} = 2\,\frac{1 + 2p}{2 + p}.$$

Hence $\rho_1 = Z_1(a_1) = 2$.

4. From (7) we get

$$A_1(0) = 2, \qquad A_2(0) = 4.$$

5. We solve the system of equations

$$A_1' - A_1^2 - 1 = 0,$$
$$A_1' A_1 + A_2 A_1' + A_2^2 - A_1^2 = 0.$$

The integral of the first of these equations is

$$A_1(\tau) = \operatorname{cth}(\tau + C_1),$$

where C_1 is the integration constant, $C_1 = \operatorname{Arc\,coth} A_1(0) \approx 0.55$.

One particular solution of the second equation is $A_2 = -1$, and therefore we seek the general solution in the form

$$A_2(\tau) = -1 + \frac{1}{\psi(\tau)}.$$

For $\psi(\tau)$ we have a first-order linear differential equation which can always be integrated in closed form

$$\psi' + \frac{2\operatorname{cth} 2(\tau + C_1)}{\operatorname{ch}(\tau + C_1)}\psi - \operatorname{th}(\tau + C_1) = 0,$$
$$\psi(\tau) = \frac{\operatorname{sh} 2(\tau + C_1) - 2(\tau + C_2)}{2\operatorname{sh} 2(\tau + C_1)},$$

where C_2 is the integration constant.

Inasmuch as $A_2(0) = 4$ we have $\psi(0) = 0.2$. On the other hand

$$\psi(0) = \frac{\operatorname{sh} 2C_1 - 2C_2}{2\operatorname{sh} 2C_1}.$$

Hence $C_2 = 0.4$.

Substituting $\psi(\tau)$ into the expression for $A_2(\tau)$ and recognizing that $W_2(\tau) = A_2^2(\tau)$, we obtain finally

$$W_2(\tau) = \left[\frac{\operatorname{sh} 2(\tau + 0{,}55) + 2(\tau + 0{,}4)}{\operatorname{sh} 2(\tau + 0{,}55) - 2(\tau + 0{,}4)}\right]^2.$$

It is appropriate to note that $W_2(\tau)$ tends monotonically to unity as $\tau \to \infty$. We can therefore always choose such $\tau = \tau_1$ at which $W_2(\tau_1)$ differs from unity by a certain specified value. If we break the line at the point $\tau = \tau_1$ and connect a load resistance equal to unity, then the required input impedance $Z_2(p)$ is ensured with prescribed accuracy by a line of finite length, loaded by a pure resistance.

2. EQUIVALENT CIRCUITS OF INHOMOGENEOUS LINES

Assume that we know the impedance of a certain network, and we denote it by Z_{k-1}. We connect this network to the output of a symmetrical two-port network with a characteristic resistance of ρ_k and a transfer function $g_k = \operatorname{Ar\,th} \frac{P}{a_k}$. Then the input impedance of the two-port network Z_k, loaded by the impedance Z_{k-1}, will be [2]

$$Z_k = \rho_k \frac{Z_{k-1} + \rho_k \operatorname{th} g_k}{\rho_k + Z_{k-1} \operatorname{th} g_k}.$$

Since the last expression does not differ at all from (2), it follows that the input impedance of a line of class k is a result of the transfer of the input impedance of a line of class k-1 by means of a two-port network with characteristic resistance ρ_k which is independent of the frequency, and a transfer function $g_k = \mathrm{Ar\,th}\,\frac{p}{a_k}$, which is independent of the coordinate τ.

Since a line of class zero is an infinitely long homogeneous line, the input impedance of an infinite line of class n can be determined from the circuit shown in Fig. 1.

Let us construct now the equivalent circuit of an inhomogeneous line of class n, loaded by an impedance Z_l. Let the delay time of the line be t_d. When $\tau = 0$ we can, by using formula (2), separate n two-port networks, just as we did for infinitely long lines.

When $\tau = t_d$ we have $Z_n(t_d) = Z_l$. At the same time we have from (2) (k = n)

$$Z_{n-1}(t_d) = \rho_n(t_d)\,\frac{Z_l - \rho_n(t_d)\,\mathrm{th}\,\frac{p}{a_n}}{\rho_n(t_d) - Z_l\,\mathrm{th}\,\frac{p}{a_n}}.$$

Here $Z_{n-1}(t_d)$ is the load resistance of the line of class n-1.

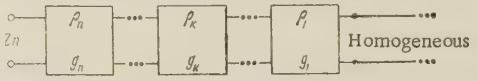


Fig. 1. Diagram for determining the input impedance of an infinitely long line of class n.

It is easy to see that $Z_{n-1}(t_d)$ is the result of transforming the impedance Z_l by means of a symmetrical two-port with characteristic resistance $\rho_n(t_d)$ and a transfer function $-g_n$.

Continuing the process of separating the two-port networks further, we obtain ultimately the equivalent network shown in Fig. 2.

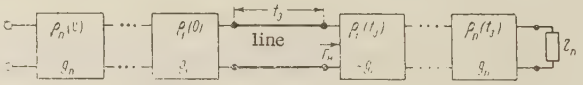
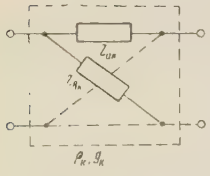


Fig. 2. Equivalent circuit of inhomogeneous line of class n loaded by an impedance Z_l.

Each of the two-ports of Fig. 2 can be replaced by a lattice, T or Pi network. In particular, for the lattice network (Fig. 3) we have



$$Z_{ak} = \frac{\rho_k}{p}\left(a_k + \sqrt{a_k^2 - p^2}\right), \quad Z_{bk} = \frac{\rho_k p}{a_k + \sqrt{a_k^2 - p^2}}.$$

Fig. 3. Replacement of a two-port with parameters ρ_k and g_k by means of a lattice network.

The equivalent circuit obtained for the inhomogeneous line enables us to take into account the influence of the load resistance Z_l on the frequency characteristics of the network.

Thus, in the synthesis of inhomogeneous line filters one usually specifies the frequency dependence of the absolute value of the reflection coefficient $\Gamma(\omega)$. It can be shown [3] that

$$\left|\frac{\Gamma_{\inf} - \Gamma_l}{1 - \Gamma_{\inf}\Gamma_l}\right| \leq |\Gamma| \leq \left|\frac{\Gamma_{\inf} + \Gamma_l}{1 + \Gamma_{\inf}\Gamma_l}\right|, \tag{9}$$

where Γ_{inf} — modulus of coefficient of reflection at the input of an infinitely long line, and Γ_l is the modulus of the coefficient of reflection at the end of a homogeneous line (Fig. 2).

If $Z_l = W_n(t_d)$ and t_d is sufficiently large, then $\Gamma_l \simeq 0$. Then $\Gamma \simeq \Gamma_{inf}$, and formula (9) enables us to estimate the error resulting in this case.

3. EXAMPLES

The method proposed, in the author's opinion, is an important means of synthesizing networks with distributed parameters. One can already point out several regions, where the use of this method will lead to important results: 1) synthesis of shaping lines; 2) synthesis of line-filters with specified working attenuation chaacteristics; 3) synthesis of microwave oscillating systems; 4) synthesis of matching junctions; 5) synthesis of transforming lines.

Since we cannot, naturally, discuss this in detail in this article, we shall consider only two very simple examples.

Example 1. Determine the variation of the wave resistance of an inhomogeneous line, which produces in a load resistance R_l, shunted with a capacitance C_l, a pulse with the following parameters: duration of pulse $t_p = 0.1$ microsecond, duration of front $t_f = 0.02$ microsecond, Load data: $R_l = 500$ ohms, $C_l = 200$ micromicrofarad.

The ratio of the voltage across Z_l to the charging voltage for the time interval $0 - t_p$ is determined by the expression

$$\frac{\bar{u}}{E} = \frac{Z_l}{Z_l + Z_1},$$

where

$$Z_l = \frac{R_l}{1 + pR_lC_l}; \qquad Z_1 = \rho_1(0)\,\frac{\frac{W_0}{\rho_1(0)} + \frac{p}{a_1}}{1 + \frac{W_0}{\rho_1(0)}\,\frac{p}{a_1}}.$$

In order for u(t) to vary exponentially (as when a homogeneous line is used) and for the pulse amplitudes to equal 0.5E, we put $W_0 = R_l$ and $a_1 = 1/\rho_1(0)C_l$. Then

$$\frac{\bar{u}}{E} = \frac{1}{2}\left(1 - \frac{p}{p + \frac{2R_l}{\rho_1^2(0)\,C_l}}\right).$$

Hence

$$t_\Phi = 0{,}5\,\frac{\rho_1^2(0)\,C_l}{R_l}.$$

The last equation enables us to determine

$$\rho_1(0) = \sqrt{2t_\Phi\,\frac{R_l}{C_l}} = 316 \text{ ohms}$$

The law of variation of the wave resistance can be determined here from (7), or else from (5) and (6). Since we have assumed $W_0 = 1$ in the derivation of (7), it is better to use equations (5) and (6).

From (6) we get

$$\rho_1(\tau) = W_0 \operatorname{th} a_1 (C_1 + \tau).$$

The integration constant is

$$C_1 = \frac{1}{a_1}\operatorname{Ar\,th}\frac{\rho_1(0)}{W_0} = \rho_1(0)\,C_l \operatorname{Ar\,th}\frac{\rho_1(0)}{R_l} = 471 \cdot 10^{-12}.$$

Calculating a_1 and substituting all the values in (5), we obtain

$$W_1(\tau) = 500 \operatorname{th}^2 (0.745 + 15.8 \cdot 10^6 \tau); \quad 0 \leqslant \tau \leqslant 0.05 \cdot 10^{-6}.$$

Example 2. Above, in Section 1, we considered an example of a synthesis of an inhomogeneous line for a specified input impedance. It is easy to check that the impedance $Z_2(p)$ considered there corresponds to the input impedance of a high-pass filter with maximally smooth characteristic of working attenuation. Actually, if we put for the internal resistance of the source $R_i = 16$, then the reflection coefficient is

$$\gamma = \frac{Z_2 - R_i}{Z_2 + R_i} = \frac{-15}{17 + 20p + 8p^2} = \frac{-15}{17 - 8\omega^2 + 20j\omega},$$

and the square of the modulus of the reflection coefficient is

$$|\gamma|^2 = \frac{1}{1 + \frac{64}{225}(\omega^2 - 1)^2}.$$

Consequently, the introduced loss factor is

$$L = \frac{1}{1 - |\gamma|^2} = 1 + \frac{225}{64(\omega^2 - 1)^2},$$

which confirms the statement made above.

With the aid of lines of the second class we can obtain simpler band pass filters with maximally smooth characteristic. Putting in (4) $a_1 = a_2 = 1$ and $\rho_1 = \rho_2 = \rho$, we get

$$Z_2 = \frac{j\left(\omega - \frac{1}{\omega}\right) + 2\rho}{j\left(\omega - \frac{1}{\omega}\right) + \frac{2}{\rho}}.$$

If we take $R_i = \rho^2$, then

$$L = 1 + 0.25\left(\rho - \frac{1}{\rho}\right)^2 \left(\frac{\omega^2 + 1}{\omega^2 - 1}\right)^2.$$

The quantity ρ determines the maximum attenuation in the rejection band of the filter.

The wave resistance of the line is determined by the method indicated above.

CONCLUSION

The procedure developed above makes it possible to synthesize inhomogeneous lines for an input impedance specified in the form of a rational-fraction function of the frequency. The method can be used for the synthesis of microwave circuits (filters, matching devices, oscillating systems) and for the synthesis of shaping and transforming lines.

The author is deeply grateful to Ya. S. Itskhoki for much valuable advice offered during the discussion of this method.

REFERENCES

1. A. L. Fel'dshtein. Certain problems in the synthesis of inhomogeneous lines. Radiotekhnika, 1958, 13, 8, 13.
2. G. I. Atabekov. Theory of linear electric networks. Soviet Radio Press, 1960.
3. Microwave Transmission Lines (MIT series) (Russian translation), vol. 2, Soviet Radio Press, 1951.

Received by editor 1 March 1961

BIHARMONIC OSCILLATIONS IN A SELF-OSCILLATOR WITH DELAY AT A NON-SYNCHRONOUS FREQUENCY RATIO

P.A. Perepelyatnik

It is shown that at a certain delay in a self-oscillator with cubic approximation of nonlinearity it is possible to generate simultaneously two frequencies, whereas for other values of the delay not one of the stationary modes is possible in the system. The resultant nonsynchronous oscillation increases without limit, and the main oscillation is damped out.

1. FUNDAMENTAL EQUATIONS

It was shown in reference [1]* that a self-oscillating system in which the nonlinearity is approximated by an incomplete cubic polynomial in the form $i_a = S_1U_g - S_3U_g^3$ can be described by the equation

$$\ddot{U} + 2\delta\dot{U} + \omega_0^2 U + 2\delta(\mu_1 - 3\mu_3 U_\tau^2) U_\tau = 0, \tag{1}$$

where $\mu_1 = S_1R$; $\mu_3 = S_3R$.

When $\tau < \tau_{cr3}$ the solution of eq. (1) in the form $U = X\cos\psi$ is likewise not unique. A more general solution of (1) must be sought in the form

$$U = \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} X_i \cos\psi_i. \tag{2}$$

We confine our analysis to the case n = 2. Then

$$U = x_1 + x_2 = X_1\cos\psi_1 + X_2\cos\psi_2. \tag{3}$$

To determine the stability limits of the single-frequency solution we first put $X_2 \ll X_1$. Then, substituting (3) in (1) after harmonic linearization, discarding terms of higher-order of smallness and the combination frequencies, we obtain two equations

$$\ddot{x}_1 + 2\delta\dot{x}_1 + \omega_0^2 x_1 + 2\delta K_1(X_1)\dot{x}_{1\tau} = 0, \tag{4a}$$

$$\ddot{x}_2 + 2\delta\dot{x}_2 + \omega_0^2 x_2 + 2\delta K_2(x_{1\tau})\dot{x}_{2\tau} = 0, \tag{4b}$$

where

$$K_1 = \mu_1 - \frac{3}{4}\mu_3 X_1^2;$$

$$K_2 = \mu_1 - 3\mu_3 X_1^2\cos^2\psi_{1\tau}.$$

Equation (4a) has already been investigated [1]. We shall now analyze Eq. (4b). It is a linear differential equation with variable coefficient and retarded argument. A solution of

---

*The present article is a continuation of [1].

this equation is facilitated by the fact that we seek the conditions under which non-synchronous oscillations set in. In this case we can expand the variable coefficient in a Fourier series and retain only the constant component. We obtain

$$K_2(X_1, 0) = \mu_1 - \frac{3}{2}\mu_3 X_1^2.$$

Equation (4b) now has the form

$$\ddot{x}_2 + 2\delta\dot{x}_2 + \omega_0^2 x_2 + 2\delta K_2(X_1, 0)\dot{x}_{2\tau} = 0. \tag{5}$$

The physical meaning of this operation is as follows. In the presence of oscillation x_1, the operating point moves along the nonlinear characteristic with frequency ω_1, and the instantaneous value of the transconductance varies in this case as $S(t) = S_1 - 3S_3X_1{}^2\cos^2\psi$. Since our system has a finite bandwidth, it will respond only to a change in the dc component of the transconductance $S_0(X_1) = S_1 - \frac{3}{2}S_3X_1^2$. The dc component, i.e., the time-averaged value of the transconductance, depends on the amplitude of the fundamental frequency oscillations. With increase in this amplitude, the value of the transconductance first decreases, and then, after passing through zero, becomes negative and then again starts increasing in absolute value.

A similar effect is sufficiently well known for the case when application of an external alternating voltage to an underexcited generator causes excitation of the natural-frequency oscillations, which are not in synchronism with the frequency of the external generator. In our case the role of the external source of alternating voltage is assumed by the generator itself.

Applying the Laplace transformation to (5) we obtain, assuming zero initial conditions

$$p^2 + 2\delta p[1 + K_2(X_1, 0)e^{-p\tau}] + \omega_0^2 = 0. \tag{6}$$

We see therefore that if $|K_2| > 1$, then there is excited in the system, at certain values of τ, a second oscillation which is not synchronous with the first.

In the general case, when x_2 is not small, we obtain in place of (4)

$$\begin{aligned} \ddot{x}_1 + 2\delta\dot{x}_1 + \omega_0^2 x_1 + 2\delta K_1(X_{1\tau}, X_{2\tau})\dot{x}_{1\tau} = 0, \\ \ddot{x}_2 + 2\delta\dot{x}_2 + \omega_0^2 x_2 + 2\delta K_2(X_{1\tau}, X_{2\tau})\dot{x}_{2\tau} = 0. \end{aligned} \tag{7}$$

Here

$$K_1 = \mu_1 - \frac{3}{4}\mu_3(X_{1\tau}^2 + 2X_{2\tau}^2);$$

$$K_2 = \mu_1 - \frac{3}{4}\mu_3(2X_{1\tau}^2 + X_{2\tau}^2);$$

$X_\tau = X$ in the case of stationary oscillations.

Taking the Laplace transformation of (7) with zero initial conditions we obtain

$$p^2 + 2\delta p(1 + K_1e^{-p\tau}) + \omega_0^2 = 0, \tag{8a}$$

$$p^2 + 2\delta p(1 + K_2e^{-p\tau}) + \omega_0^2 = 0. \tag{8b}$$

Substituting $p = j\omega_1$ in (8a) and $p = j\omega_2$ in (8b) and equating the real and imaginary parts of the resultant expressions to zero, we obtain

$$\begin{aligned} 1 + K_i\cos\omega_i\tau = 0, \\ \omega_0^2 - \omega_i^2 + 2\delta K_i\omega_i\sin\omega_i\tau = 0. \end{aligned} \qquad i = 1, 2 \tag{9}$$

Expressions (9) can be written in different form:

$$\frac{\omega_0}{\omega_i} - \frac{\omega_i}{\omega_0} = \frac{1}{Q}\sqrt{K_i^2 - 1},$$
$$\operatorname{tg}\omega_i\tau = Q\left(\frac{\omega_0}{\omega_i} - \frac{\omega_i}{\omega_0}\right), \tag{10}$$

where $Q = R\omega_0 C$.

After suitable transformation Eq. (10) can be represented as

$$W_1 = \frac{K_1}{\sqrt{1 + Q^2\left(\frac{\omega_0}{\omega_1} - \frac{\omega_1}{\omega_0}\right)^2}} = 1, \tag{11a}$$

$$\tau = \frac{1}{\omega_1}\left[\operatorname{arc\,tg} Q\left(\frac{\omega_0}{\omega_1} - \frac{\omega_1}{\omega_0}\right) + (2n+1)\pi\right];\quad n = 0, 1, 2, \ldots, \tag{11b}$$

$$W_2 = \frac{|K_2|}{\sqrt{1 + Q^2\left(\frac{\omega_0}{\omega_2} - \frac{\omega_2}{\omega_0}\right)^2}} = 1, \tag{12a}$$

$$\tau = \frac{1}{\omega_2}\left[\operatorname{arc\,tg} Q\left(\frac{\omega_0}{\omega_2} - \frac{\omega_2}{\omega_0}\right) + (2m+1)\pi\right];\quad m = 0, 1, 2, \ldots \text{ for } K_2 \geqslant 1, \tag{12b}$$

$$\tau = \frac{1}{\omega_2}\left[\operatorname{arc\,tg}\, Q\left(\frac{\omega_0}{\omega_2} - \frac{\omega_2}{\omega_0}\right) + 2m\pi\right];\quad m = 0, 1, 2, \ldots \text{ for } K_2 \leqslant -1. \tag{12c}$$

In (11) we assume that K_1 is always positive. In (12c) with m = 0 we take only those values of the frequencies, for which τ is positive.

As already indicated, non-synchronous oscillations cannot occur when $|K_2| < 1$. We are therefore interested only in that region of the fundamental frequency ω_1, for which $|K_2| \geq 1$.

If conditions (11) and (12) are simultaneously satisfied for some value of the delay time, then non-synchronous oscillations can exist in the system.

We shall call that minimum value of the delay time, for which two non-synchronous oscillations can exist the fourth critical delay time (τ_{cr4}).

Using (11) and (12) we can find that in the case of $K_2 < -1$

$$\tau_{cr4} = \frac{\pi - (|\varphi_1| + |\varphi_2|)}{|\omega_2 - \omega_1|}, \tag{13}$$

where $\varphi_i = \operatorname{arc\,tg} Q\left(\frac{\omega_0}{\omega_i} - \frac{\omega_i}{\omega_0}\right)$; $i = 1, 2$. In the case $K_2 > 1$ we have $\tau_{cr4} = \tau_{cr2}$.

To determine the stability of the resultant non-synchronous oscillations we must make a special investigation.

2. INVESTIGATION OF THE STABILITY OF NON-SYNCHRONOUS OSCILLATIONS

We put

$$x_i = (X_{i0} + \xi_i)\cos(\psi_{i0} + \theta_i);\quad i = 1, 2, \tag{14}$$

where ξ_i and θ_i are small, slowly varying functions of the time

Expanding (14) in a Taylor series in the vicinity of X_{i0} and ψ_{i0}, we obtain for the principal part of the increment

$$dx_i = \cos\psi_{i0}\,\xi_i - X_i \sin\psi_{i0}\theta_i;\quad i = 1, 2. \tag{15}$$

Using (15) and (7) we get

$$d\ddot{x}_i + 2\delta d\dot{x}_i + \omega_0^2 dx_i + 2\delta d\,(K_{i\tau}\dot{x}_{i\tau}) = 0;\quad i = 1, 2. \tag{16}$$

Discarding terms of higher order of smallness, we obtain after suitable transformations from (16)

$$-\dot{\xi}_1 T_k = \xi_1 + K_{\xi 1}\cos\omega_1\tau\xi_{1\tau} - K_{21\xi}\cos\omega_1\tau\xi_{2\tau} - K_1\sin\omega_1\tau X_{10}(\theta_1 - \theta_{1\tau}),$$
$$-\dot{\xi}_2 T_k = \xi_2 + K_{\xi 2}\cos\omega_2\tau\xi_{2\tau} - K_{21\xi}\cos\omega_2\tau\xi_{1\tau} - K_2\sin\omega_2\tau X_{20}(\theta_2 - \theta_{2\tau}),$$

$$\begin{aligned}-\dot{\theta}_1 T_k &= \frac{1}{X_{10}}\sin\omega_1\tau\xi_1 - \frac{1}{X_{10}}\sin\omega_1\tau K_{\xi 1}\xi_{1\tau} + \\ &+ \frac{1}{X_{10}}\sin\omega_1\tau K_{21\xi}\xi_{2\tau} + \theta_1 + \cos\omega_1\tau K_1\theta_{1\tau},\\ -\dot{\theta}_2 T_k &= \frac{1}{X_{20}}\sin\omega_2\tau\xi_2 - \frac{1}{X_{20}}\sin\omega_2\tau K_{\xi 2}\xi_{2\tau} + \\ &+ \frac{1}{X_{20}}\sin\omega_2\tau K_{21\xi}\xi_{1\tau} + \theta_2 + \cos\omega_2\tau K_2\theta_{2\tau}.\end{aligned} \tag{17}$$

We choose here the notation

$$K_{\xi 1} = \mu_1 - \frac{9}{4}\mu_3\left(X_{10}^2 + \frac{2}{3}X_{20}^2\right);$$
$$K_{\xi 2} = \mu_1 - \frac{9}{4}\mu_3\left(X_{20}^2 + \frac{2}{3}X_{10}^2\right);$$
$$K_{21\xi} = 3\mu_3 X_{10}X_{20}.$$

Changing to the Laplace transform and being interested only in the stability, we can obtain from (17) the characteristic equation

$$\begin{vmatrix} a_{11}+\lambda & a_{12} & a_{13} & 0 \\ b_{21} & b_{22}+\lambda & 0 & b_{24} \\ c_{31} & c_{32} & c_{33}+\lambda & 0 \\ d_{41} & d_{42} & 0 & d_{44}+\lambda \end{vmatrix} = 0. \tag{18}$$

Here

$$a_{11} = 1 - \frac{K_{\xi 1}}{K_1}e^{-p\tau}; \quad a_{12} = \frac{K_{21\xi}}{K_1}e^{-p\tau};$$
$$a_{13} = -K_1\sin\omega_1\tau X_{10}(1 - e^{-p\tau}); \quad b_{21} = \frac{K_{21\xi}}{K_2}e^{-p\tau};$$
$$b_{22} = 1 - \frac{K_{\xi 2}}{K_2}e^{-p\tau}; \quad b_{24} = -K_2\sin\omega_2\tau X_{20}(1 - e^{-p\tau});$$
$$c_{31} = \frac{1}{X_{10}}\sin\omega_1\tau(1 - K_{\xi 1}e^{-p\tau}); \quad c_{32} = \frac{1}{X_{10}}\sin\omega_1\tau K_{21\xi}e^{-p\tau};$$
$$c_{33} = 1 - e^{-p\tau}; \quad d_{41} = \frac{1}{X_{20}}\sin\omega_2\tau K_{21\xi}e^{-p\tau};$$
$$d_{42} = \frac{1}{X_{20}}\sin\omega_2\tau(1 - K_{\xi 2}e^{-p\tau});$$
$$d_{44} = c_{33}; \qquad \lambda = pT_k.$$

It was shown in [1] that when $\tau > T_{cr3}$, the single-frequency self-oscillations are unstable, since self modulation is produced. The self-modulation can be either hard, if sufficiently large disturbances are necessary, or soft. In this case the self modulation occurs spontaneously.

From physical considerations it is quite obvious that in the case of two-frequency oscillations at large delays the oscillations will also be unstable. Therefore there is no need to investigate the characteristic equation (18) in the general case. We shall consider the possibility of simultaneous generation at $p\tau \simeq 0$, when it is possible to put $|e^{-p\tau}| \simeq 1$. In this case the characteristic equation (18) simplifies to

$$\lambda^2 + \lambda (a_{11} + b_{22}) + a_{11}b_{22} - a_{12}b_{21} = 0. \tag{19}$$

The non-synchronous oscillations will be stable if the following inequalities are simultaneously satisfied

$$a_{11}b_{22} - a_{12}b_{21} > 0, \tag{20}$$

$$a_{11} + b_{22} > 0. \tag{21}$$

If $K_1 \geq 1$ and $K_2 > 1$, condition (20) is not satisfied, in agreement with the results of [2] and [3]. Physically this means that the main oscillations, causing the non-synchronous oscillation, will be subsequently suppressed by the non-synchronous oscillation itself, and the system will change over to generate only one new frequency.

If $K_1 \geq 1$ and $K_2 \leq 1$, then the inequality (20) is fulfilled. In this case, for the non-synchronous oscillations to be stable it is necessary to satisfy the additional condition (21), which can be converted to the form

$$\frac{X_{10}^2}{K_1} - \frac{X_{20}^2}{|K_2|} > 0. \tag{22}$$

This still does not complete the analysis of the stability of the biharmonic oscillations. It is necessary to investigate in addition the possibility of occurrence of a third oscillation. Let us put

$$U = x_1 + x_2 + x_3 = X_1 \cos \psi_1 + X_2 \cos \psi_2 + X_3 \cos \psi_3, \tag{23}$$

with $X_3 \ll X_1$ and X_2. Substituting (23) into the fundamental equation (1) we obtain after suitable transformations the equations (7), which we have already investigated, and for the oscillation X_3 we obtain the equation

$$\ddot{x}_3 + 2\delta x_3 + \omega_0^2 x_3 + 2\delta [\mu_1 - 3\mu_3 (x_{1\tau}^2 + x_{2\tau}^2)] \dot{x}_{3\tau} = 0. \tag{24}$$

Equation (24) is also a linear differential equation with variable coefficient and retarded argument. Since we are interested in the non-synchronous oscillations, we shall expand the variable coefficient in a double Fourier series and retain only the constant coefficient, thereby obtaining

$$\ddot{x}_3 + 2\delta \dot{x}_3 + \omega_0^2 x_3 + 2\delta K_3 (X_1, X_2, 0) \dot{x}_{3\tau} = 0, \tag{25}$$

where

$$K_3 = \mu_1 - \frac{3}{2} \mu_3 (X_1^2 + X_2^2).$$

Applying the Laplace transformation (25) and assuming zero initial conditons, we obtain

$$p^2 + 2\delta p (1 + K_3 e^{-p\tau}) + \omega_0^2 = 0. \tag{26}$$

If $|K_3| > 1$, then oscillations at a third frequency can occur for a certain value of τ. In other words, if for a given delay there is a frequency ω_3, located in an interval that includes ω_1 and ω_2, at which the following conditions are satisfied

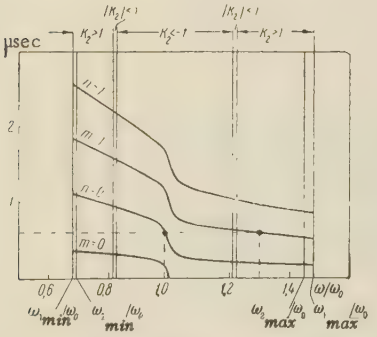
$$\tau = \frac{1}{\omega_3} \left[\operatorname{arc\,tg} Q \left(\frac{\omega_0}{\omega_3} - \frac{\omega_3}{\omega_0} \right) + 2m\pi \right]; \quad m = 0, 1, 2, \ldots, \quad K_3 < -1, \tag{27}$$

$$W_3 (X_1, X_2, 0) = \frac{|K_3 (X_1, X_2, 0)|}{\sqrt{1 + Q^2 \left(\frac{\omega_0}{\omega_3} - \frac{\omega_3}{\omega_0} \right)^2}} > 1, \tag{28}$$

then the amplitude of the oscillations of frequency ω_3 will start increasing. In this case the biharmonic mode under consideration, with frequencies ω_1 and ω_2, will be unstable.

In equation (28) W_3 is the closed-loop transfer coefficient for the oscillations of frequency ω_3 in the presence of the two investigated non-synchronous oscillations.

To answer the question of how the system will behave in this case, it is necessary to consider the possibility of simultaneous generation of three frequencies.



Only if one of the conditions (27) or (28) is not satisfied, will the biharmonic mode be stable. If condition (22) is not satisfied, then stable oscillations cannot exist in the system at a given value of delay. In this case the resultant oscillation with frequency ω_2 suppresses the fundamental oscillation frequency ω_1 and then increases without limit. The system considered will not have a single stable solution at this value of delay. Such a case is possible only when the value of the non-synchronous frequency ω_2 is sufficiently close to the resonant frequency ω_0 of the tuned circuit.

In a physical system the oscillations cannot go to infinity, for there is always saturation. After all, any physical system has a power limit beyond which it cannot supply the load.

We have obtained a result that contradicts this conclusion only because we have mathematically idealized the nonlinearities, without regard for the physical system.

In addition to the non-synchronous mode, synchronous modes are also possible in the system, in which the generated frequencies are related as $n\omega_1 = m\omega_2$, where n and m are integers. Generally speaking, if $n + m \gg 1$, then the oscillations can be regarded as non-synchronous; if n and m are small integers, the oscillations will be synchronous and not only their amplitudes but also their phases will be interrelated.

An investigation of the synchronous oscillations entails greater difficulties and is not considered in the present paper.

For the sake of illustration of the resultant equations, let us consider the following example. Let $S = 10^{-3} - 10^{-5} X^2$, $Q = 50$, $R = 40$ kilohms and $\omega_0 = 2\pi \cdot 10^6$ 1/sec. The figure shows the dependence of the delay time on the frequency, when equations (11b) and (12c) are satisfied; $\omega_{1\min}$ and $\omega_{1\max}$ indicate the bandwidth of the system in the case when only one frequency is generated; $\omega_{2\min}$ and $\omega_{2\max}$ indicate the bandwidth for the non-synchronous frequency, if the fundamental frequency coincides with the resonant frequency of the tank circuits. If the fundamental frequency does not coincide with the resonant frequency of the tank circuit, then the transmission band for the non-synchronous oscillation will be smaller. The figure indicates the regions where the transfer coefficient of the first harmonic of the non-synchronous oscillation has different values. The case when $\omega_1 = \omega_0$ is most favorable for the occurrence of non-synchronous oscillations.

When $\tau = 0.6$ microsecond, two stable non-synchronous oscillations are possible with frequencies $\omega_1 = 0.99\ \omega_0$; $\omega_2 = 1.3\ \omega_0$ and respective amplitudes $X_1 = 10.3$ v and $X_2 = 3.5$ v.

If $\tau = 0.8$ microsecond, not one stable oscillation can exist in the system.

CONCLUSION

We have shown that in a self-oscillator with delay and with nonlinearity in the form of an incomplete cubic polynomial, under certain values of the delay, greater than or equal to the fourth critical value (τ_{cr4}), there can exist a stable biharmonic mode. We found an expression for the fourth critical value of the delay. We have shown that there exist such values of the delay ($\tau \geq \tau_{cr4}$) at which there are no stable oscillations in the system at all.

REFERENCES

1. P.A. Perepelyatnik. Self-oscillation in a generator with delay. Radiotekhnika i elektronika 1961, 6, 10, 1601.
2. V.S. Pankratov. Stationary modes of a self-oscillator with delayed feedback. Izv. vuzov MVO SSSR, Radiotekhnika, 1958, 1, 6, 705.
3. V. Met. On multimode oscillators with constant time delay, ProcIRE 1957, 45, 8, 1119.

Received by editor 19 November 1960

CHAIN SYNTHESIS OF A REACTANCE MULTI-PORT NETWORK

A.I. Rutkas

Certain new mathematical researches have led to a method for representing multi-port LC networks in the form of an equivalent chain of simpler circuits. The method is illustrated by a numerical example.

INTRODUCTION

An important tool in waveguide theory is the use of an electric circuit that is equivalent to the inhomogeneous part of the waveguide. To study the character of propagation of the waves through the inhomogeneity it is interesting to construct such a circuit in the form of a transmission line with lumped or distributed parameters L and C, the parameters being independent of the frequency. A solution of this problem can be used for synthesis of a waveguide to satisfy certain frequency characteristics.

In this connection, the question arises of replacing the multi-port network by an equivalent chain of simpler circuits.

We consider below passive reactance for 2m-ports with lumped parameters L and C, operating in steady state and connected to an external electric system such as to ensure equality of the input and output currents of each pair of terminals [5, 9]. It turns out that the transfer matrix $A(\omega)$ of such a 2m-port (the analog of the matrix of generalized parameters of the two-port network) has the property that it does not stretch a certain indefinite metric J in the corresponding linear space for any frequency ω with $\operatorname{Im}\omega \geq 0$. Recently there have been developed general mathematical methods of expanding analytic matrices of this type into a product of simpler matrix multipliers [1, 2, 3].

We propose here to use these methods for realization of a matrix $A(\omega)$ in the form of a chain of simpler 2m-ports. We note that when so connected their transfer matrices are multiplied in reverse order. As applied to two-port networks, this idea was proposed and realized in [4].*

The general method considered for factoring the analytic matrix functions can also be used as the basis of synthesis of a multi-conducted transmission line with distributed parameters L and C.

1. THE MATRIX $A(\omega)$

Leaving aside cases of degeneracy, we assume that the currents and voltages at the input of the 2m-port are linearly independent (we refer to the complex amplitude values of the currents and voltages).

In n-dimensional linear space H (n = 2m) we shall call the transfer matrix $A(\omega)$ a matrix-function of the frequency ω, expressing the vector of the state at the output $\vec{X}'$ in terms of the vector of the state at the input $\vec{X}$:

---

*This problem was also solved for a reactance two-port by a different method in [8].

$$\vec{X}' \quad A(\omega)\vec{X}; \quad X' = \begin{pmatrix} U_1' \\ \vdots \\ U_m' \\ I_m' \\ \vdots \\ I_1' \end{pmatrix}; \qquad X = \begin{pmatrix} U_1 \\ \vdots \\ U_m \\ I_m \\ \vdots \\ I_1 \end{pmatrix}. \tag{1}$$

For 2m-ports of this type, $A(\omega)$ exists by virtue of the linearity of the connection between the currents and the voltages applied to the terminals. We note that the elements of $A(\omega)$ can be expressed in terms of the elements of the S and T wave matrices of the 2m-ports [6, 7].

Assume that in some orthonormalized basis of the space H the matrix $A(\omega)$ specifies an operator $A(\omega)$. We define in this basis, with the aid of the matrix whose elements on the lateral diagonal are unity and in all other places zero, the operator $J(J = J^*, J^2 = 1)$. The operator J introduces in H an indefinite metric with J-scalar product $[\vec{X}, \vec{Y}] = (J\vec{X}, \vec{Y})$, where the round brackets denote the ordinary scalar product. The operator T is called J-nonstretching, if for any $\vec{X} \in H$ $[T\vec{X}, T\vec{X}] \leq [\vec{X}, \vec{X}]$.

Using (1) to obtain $J\vec{X}$ and $J\vec{X}'$, we get

$$[A\vec{X}, A\vec{X}] = [\vec{X}', \vec{X}'] = (J\vec{X}', \vec{X}') = I_1'U_1'^* + \ldots + I_m'U_m'^* + U_m'I_m'^* + \ldots + U_1'I_1'^* = 2\mathrm{Re}\,(U_1'I_1'^* + \ldots + U_m'I_m'^*),$$
$$[\vec{X}, \vec{X}] = 2\mathrm{Re}\,(U_1I_1^* + \ldots + U_mI_m^*).$$

The effective power applied to the terminals of the 2m-port is

$$W_{\text{ef}} = \tfrac{1}{2}\mathrm{Re}\,(U_1I_1^* + \ldots + U_mI_m^*) - \tfrac{1}{2}\mathrm{Re}\,(U_1'I_1' + \ldots + U_m'I_m'^*) \geqslant 0$$

where $\mathrm{Im}\omega \geq 0$ (see, for example, [5]). It follows therefore that $[A\vec{X}, A\vec{X}] \leq [\vec{X}, \vec{X}]$, and the and the equality takes place if and only if $\mathrm{Im}\omega = 0$.

Thus, $A(\omega)$ is an operator function which is J-nonstretching in the upper halfplane and J-unitary on the real ω axis. This property can be written in the form

$$J - A^*JA \geqslant 0, \quad \mathrm{Im}\,\omega \geqslant 0, \tag{2}$$

since $(JA\vec{X}, A\vec{X}) \leq (J\vec{X}, \vec{X})$ implies $(J\vec{X}, \vec{X}) - (A^*JA\vec{X}, \vec{X}) \geq 0$.

The elements of the transfer matrix are analytic (rational) functions of ω; a similar operator function $A(\omega)$ is called analytic (in particular, rational).

Two multi-port networks will be called equivalent if their transfer matrices coincide (after W. Cauer [9]).

2. EXPANSION OF $A(\omega)$ INTO A PRODUCT

For an operator-function having property (2), this question has been considered in [1, 2, 3]. We are faced with the problem of synthesizing a matrix $A(\omega)$ in the form of a chain of simpler 2m-ports; it is therefore necessary to expand the operator function $A(\omega)$ into a product of elementary factors. By the methods developed in the indicated papers, we can obtain a solution of this problem, which we formulate in the form of the following fundamental theorem (without proof).

The transfer matrix $A(\omega)$ can be represented as a finite product

$$A(\omega) = UB_n(\omega) \ldots B_2(\omega)B_1(\omega), \tag{3}$$

where U is a J-unitary operator independent of ω; $B_j(\omega)$ is an operator-function having property (2) and given by the formula

$$B_j(\omega) = I - \frac{i}{\omega - \omega_0} Q_{\alpha_j}, \tag{4}$$

in which $Q_{\alpha i}$ has the form I or II:

$$Q_{\alpha_i} = -2\mathrm{Im}\ \omega_0 R_{\alpha_i} = -2\mathrm{Im}\ \omega_0\, JC^*\,(P_{\alpha_i} CJC^*)^{[-1]} C. \tag{5}$$

Here $R^2_{\alpha_i} = R_{\alpha i}$ — highest-order coefficient of the Laurent expansion in the vicinity of the pole ω_0 ($\mathrm{Im}\ \omega_0 \neq 0$) of the operator function $A_{j-1}(\omega) = A(\omega)\, B_1^{-1}(\omega)$; $CJC^* < 0$; α_i — one of the eigenvalues of the self-adjoint operator CJC^*; P_{α_i} — operator of projection on a one-dimensional invariant subspace of the transformation CJC^*, corresponding to the eigenvalue α_i.

By the operation $D^{[-1]}$ we mean the inversion of the operator, induced by the self-adjoint operator D on its own domain of values.

$$Q_{\alpha_i} = JC^*\,[P_{\alpha_i}(-iDJC^*)]^{[-1]} C, \quad Q^2_{\alpha_i} = 0. \tag{6}$$

Here $A_{j-1}(\omega) = (\omega - \omega_0)^{-k} C + (\omega - \omega_0)^{-k+1} D + \ldots$ ($\mathrm{Im}\,\omega_0 = 0$), and $C \neq 0$, $CJC^* = 0$; $DJC^* = -iDJC^* > 0$. The expression $[P_{\alpha_i} \times (-iDJC^*)]^{[-1]}$ is defined in analogy with the preceding one.

3. CASE WHEN A(ω) HAS POLES AT 0 AND ∞

Let us consider the important case when the elements of $A(\omega)$ have the following form: $a_{ik}(\omega) = P_{ik}(\omega) + q_{ik}\left(\frac{1}{\omega}\right)$; where p_{ik}, q_{ik} are certain polynomials.

It can be shown that in this case each of the factors $B_j(\omega)$ in product (3) is physically realizable. For the sake of being specific we shall put henceforth n = 4 and m = 2 (four-port network). In this case $B_j(\omega)$ will be a matrix-function of one of the following four types:

$$I.\quad B_j(\omega) = I - \frac{i}{\omega}\left(\begin{array}{c|c} 0 & q \\ \hline 0 & 0 \end{array}\right),$$

$$II.\quad B_j(\omega) = I - \frac{i}{\omega}\left(\begin{array}{c|c} 0 & 0 \\ \hline q & 0 \end{array}\right),$$

$$III.\quad B_j(\omega) = I + i\omega\left(\begin{array}{c|c} 0 & q \\ \hline 0 & 0 \end{array}\right),$$

$$IV.\quad B_j(\omega) = I + i\omega\left(\begin{array}{c|c} 0 & 0 \\ \hline q & 0 \end{array}\right),$$

Type	I	II	III	IV
L	—	$1/\rho_3$	ρ_3	—
C	$1/\rho_3$	—	—	ρ_3
K	ρ_2/ρ_1	$-\rho_2/\rho_1$	ρ_2/ρ_1	$-\rho_2/\rho_1$

where $q = \begin{pmatrix} \rho_1 & \rho_2 \\ \rho_3 & \rho_1 \end{pmatrix}$; ρ_1, ρ_2, ρ_3 — real numbers, such that $\rho_2 \geq 0$, $\rho_3 \geq 0$, $\rho_1^2 = \rho_2\rho_3$.

Regarding $B_j(\omega)$ as transfer matrices, we obtain their realization respectively in the form of the four-ports I — IV (Fig. 1). The values of K, L, and C are listed in the table.

If any of the coefficients ρ_1, ρ_2, ρ_3 vanishes, then the aforementioned four-ports degenerate into two two-ports 121'2' and 343'4' which are not coupled with each other.

CONCLUSION

Any four-port (or accordingly any 2m-port) with transfer matrix of the type indicated can be represented by an equivalent chain of irreducible 4-ports of the form I — IV (Fig. 1).

To effect chain synthesis in the case of an arbitrary transfer matrix $A(\omega)$, it is necessary to unify in groups of two or in groups of four the factors $B_j(\omega)$ corresponding to those poles of $A(\omega)$, which are symmetrical about the coordinate axes. In this case the results established in Section 3 remain in force for the factors $B_j(\omega)$ corresponding to $\omega_0 = 0$ and $\omega_0 = \infty$.

I am sincerely grateful to M.S. Livshitz for interest in the work.

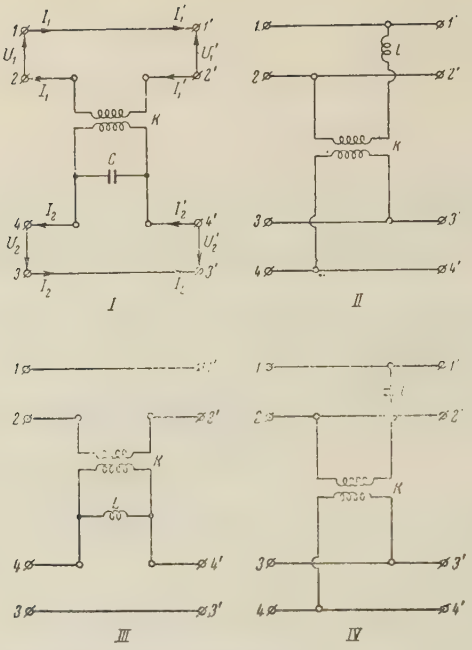


Figure 1

Example

Let a certain four-port have a transfer matrix

$$A(\omega)=\left(\begin{array}{cc|cc} 1 & 0 & \frac{1}{\omega}(-2i) & \frac{1}{\omega}(-3i) \\ 0 & 1 & \frac{1}{\omega}(-12i) & \frac{1}{\omega}(-2i) \\ \hline \frac{1}{\omega}(3i) & \frac{1}{\omega}(-i) & 1+\frac{1}{\omega^2}(-6) & \frac{1}{\omega^2}(7) \\ \frac{1}{\omega}(-9i) & \frac{1}{\omega}(3i) & \frac{1}{\omega^2}(18) & 1+\frac{1}{\omega^2}(-21) \end{array}\right)=\left(\begin{array}{c|c} I & A_b \\ \hline A_c & A_d \end{array}\right).$$

Let us find the equivalent chain of elementary four-ports I — IV.

$A(\omega)$ has one second-order pole at $\omega_0 = 0$.

$$1.\ A(\omega)=A_1(\omega)\,B_1(\omega);\quad B_1(\omega)=I-\frac{i}{\omega}Q_1,$$

where Q_1 is given by (6).

$$C=\left(\begin{array}{c|cc} 0 & 0 & \\ \hline 0 & -6 & 7 \\ & 18 & -21 \end{array}\right),\qquad D=\left(\begin{array}{cc|cc} & 0 & -2i & -3i \\ & & -12i & -2i \\ \hline 3i & -i & & \\ -9i & 3i & 0 & \end{array}\right),$$

$$K_f=-iDJC^*=\left(\begin{array}{c|c} 0 & 0 \\ \hline 0 & R_f \end{array}\right),\qquad R_f=\begin{pmatrix} 27 & -81 \\ -81 & 243 \end{pmatrix}=\begin{pmatrix} r_{11} & r_{12} \\ r_{12} & r_{22} \end{pmatrix}.$$

Inasmuch as Det $R_f = 0$, K_f is of the first rank and $P_{\alpha 1}(-iDJC^*) = K_f$. To reduce K_f to diagonal form it is sufficient to reduce R_f to diagonal form, for which purpose we go over in H space from the initial orthonormal basis $\{f\}$ to a different orthonormal basis $\{e\}$ in which

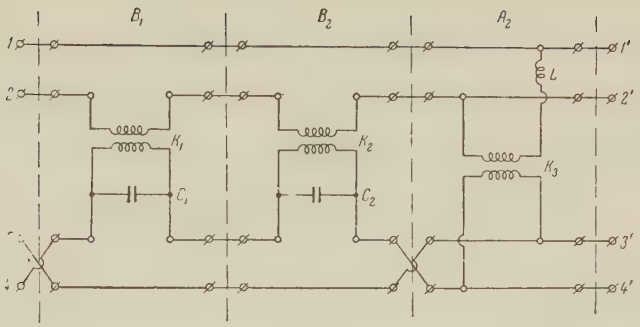


Fig. 2.

the matrix of the operator — iDJC* will have the form

$$K_e = \left(\begin{array}{c|c} 0 & 0 \\ \hline 0 & R_e \end{array}\right), \quad R_e = \begin{pmatrix} \alpha_1 & 0 \\ 0 & 0 \end{pmatrix};$$

$$\alpha_1 = 27 + 243 = 270$$

(α_1 — trace of the operator R).

Let the transition from {f} to {e} in the two-dimensional space corresponding to the operator R be given by the unitary matrix.

$$U_2 = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix}.$$

Then

$$R_f = U_2 R_e U_2' = \begin{pmatrix} \alpha_1 u_{11}^2 & \alpha_1 u_{11} u_{21} \\ \alpha_1 u_{11} u_{21} & \alpha_1 u_{21}^2 \end{pmatrix},$$

$$u_{11} = \sqrt{\frac{r_{11}}{\alpha_1}}, \quad u_{21} = \sqrt{\frac{r_{22}}{\alpha_1}};$$

$$K_f^{[-1]} = \left(\begin{array}{c|c} 0 & 0 \\ \hline 0 & R_f^{[-1]} \end{array}\right),$$

$$R_f^{[-1]} = U_2 R_e^{[-1]} U_2' = U_2 \begin{pmatrix} \frac{1}{\alpha_1} & 0 \\ 0 & 0 \end{pmatrix} \qquad U_2' = \begin{pmatrix} \frac{1}{2700} & -\frac{1}{900} \\ -\frac{1}{900} & \frac{1}{300} \end{pmatrix}.$$

Calculating now $Q_1 = JC^* K_1^{[-1]} C$, we find

$$B_1(\omega) = I - \frac{i}{\omega}\left(\begin{array}{c|c} 0 & q_1 \\ \hline 0 & 0 \end{array}\right), \quad q_1 = \begin{pmatrix} -\frac{14}{9} & \frac{49}{27} \\ \frac{4}{3} & -\frac{14}{9} \end{pmatrix}.$$

Noting that $B_1^{-1}(\omega) = I + \frac{i}{\omega} Q_1$ we obtain by multiplication

$$A_1(\omega) = A(\omega) B_1^{-1}(\omega) = \left(\begin{array}{c|c} I & A_1^b \\ \hline A_c & I \end{array}\right), \qquad A_1^b = \begin{pmatrix} \frac{1}{\omega}\left(-\frac{32}{9} i\right) & \frac{1}{\omega}\left(-\frac{32}{27} i\right) \\ \frac{1}{\omega}\left(-\frac{32}{3} i\right) & \frac{1}{\omega}\left(-\frac{32}{9} i\right) \end{pmatrix},$$

$$C_1 = \left(\begin{array}{cc|cc} & 0 & -\frac{32}{9}i & -\frac{32}{27}i \\ & & -\frac{32}{3}i & -\frac{32}{9}i \\ \hline 3i & -i & \multicolumn{2}{c}{0} \\ -9i & 3i & & \end{array}\right),$$

$$D_1 = I, \quad -iD_1JC_1^* = \left(\begin{array}{cc|cc} \frac{32}{27} & \frac{32}{9} & \multicolumn{2}{c}{0} \\ \frac{32}{9} & \frac{32}{3} & & \\ \hline & 0 & 1 & -3 \\ & & -3 & 9 \end{array}\right) = \left[\begin{array}{c|c} R_{1f} & 0 \\ \hline 0 & R_{2f} \end{array}\right],$$

R_{1f} and R_{2f} are first-rank matrices. In some new basis $\{e\}$ we have

$$-iD_1JC_1^* = \left(\begin{array}{c|c} R_{1e} & 0 \\ \hline 0 & R_{2e} \end{array}\right), \quad R_{1e} = \begin{pmatrix} \alpha_1 & 0 \\ 0 & 0 \end{pmatrix}, \quad R_{2e} = \begin{pmatrix} \alpha_2 & 0 \\ 0 & 0 \end{pmatrix}, \quad \alpha_1 = \frac{320}{27},$$

$$K_e' = P_{\alpha_1}(-iD_1JC_1^*) = \left(\begin{array}{c|c} R_{1e} & 0 \\ \hline 0 & 0 \end{array}\right).$$

In the basis $\{f\}$ we have

$$K_f^{[-1]} = \left(\begin{array}{c|c} R_{1f}^{[-1]} & 0 \\ \hline 0 & 0 \end{array}\right).$$

Using the results of the preceeding subsection, we have

$$R_{1f}^{[-1]} = \begin{pmatrix} \frac{27}{3200} & \frac{81}{3200} \\ \frac{81}{3200} & \frac{243}{3200} \end{pmatrix}, \qquad B_2(\omega) = I - \frac{i}{\omega} Q_2,$$

$$Q_2 = JC_1^* K_f^{[-1]} C_1,$$

$$Q_2 = \left(\begin{array}{c|c} 0 & q_2 \\ \hline 0 & 0 \end{array}\right), \quad q_2 = \begin{pmatrix} \frac{32}{9} & \frac{32}{27} \\ \frac{32}{3} & \frac{32}{9} \end{pmatrix}, \quad A_1(\omega) = A_2(\omega) B_2(\omega),$$

$$A_2(\omega) = A_1(\omega) B_2^{-1}(\omega) = I - \frac{i}{\omega}\left(\begin{array}{c|c} 0 & 0 \\ \hline q_3 & 0 \end{array}\right), \quad q_3 = \begin{pmatrix} -3 & 1 \\ 9 & -3 \end{pmatrix},$$

$$A(\omega) = A_2(\omega) B_2(\omega) B_1(\omega).$$

The matrices $B_1(\omega)$ and $B_2(\omega)$ are of type I, while $A_2(\omega)$ is of type II [see Section 3, p. 1841 (Russian)]. This completes the factoring of $A(\omega)$. The sought chain of 4-ports is shown in Fig. 2. From the properties of the product (3) we find that in our example all three 4-ports commute with each other.

By virtue of the relations in the paper, we have $K_1 = -7/6$; $K_2 = 1/3$; $K_3 = 1/3$; $C_1 = 3/4$; $C_2 = 3/32$; $L = 1/9$.

REFERENCES

1. M.S. Brodskii and M.S. Livshitz. Spectral analysis of non self adjoint operators and intermediate systems. Advances in Mathematical Sciences, 1958, 13, 1 (79), 3.
2. V.P. Potapov. Multiplicative structure of J-non-stretching matrix functions. Transactions of the Moscow Mathematical Society, 1954, 4, 125.
3. Yu.P. Ginzburg. On J-nonstretching operator functions. Proceedings, Academy of Sciences USSR), 1957, 117, 2, 171.
4. M.S. Livshitz, E.S. Flekser. Expansion of a reactance two-port in a chain of simpler two-ports, ibid, 1960, 135, 3, 542.
5. V.A. Taft. Fundamental procedures in the design of linear electric circuits for specified frequency characteristics, Press of the Academy of Sciences USSR, 1954.

6. L.R. Yavich. Wave matrices of a two-port network, Radiotekhnika i elektronika 1957, 2, 7, 870.
7. R.A. Silin. Properties of wave matrices of a 2(P → 1)-pole network, Radiotekhnika i elektronika, 1954, 4, 10, 1628.
8. A. Talbot. New method of synthesis of reactance networks. Proc IEE, 1954, 101, pt. 4, Monograph no. 77, p. 73.
9. W. Cauer. Theorie der linearen Wechselstromschaltungen. Akademie-Verlag, Berlin, 1954.

Received by editor 3 March 1961

CORRELATION CHARACTERISTICS OF THE FIELD OF A LINEAR ANTENNA

Ya.S. Shifrin

We consider the correlation characteristics of the field of a linear in-phase antenna in the presence of random phase errors in the sources. The correlation function obtained is used to determine the functional of the distribution of the envelope of the directivity pattern of the antenna.

INTRODUCTION

In the papers published up to now on statistical estimates of the parameters of antenna systems [1 — 4], principal attention was paid to the determination of the average characteristics (with respect to time or with respect to a family of antennas of a single type) of the directivity and the average values of certain antenna parameters, particularly the directivity coefficient.

The next step towards a study of the statistical properties of the field of an antenna is, naturally, an investigation of the fluctuations of various parameters. This question was considered in part in [3]. There exist, however, many problems of practical importance, the solution of which necessitates a deeper study of the statistics of antennas. Such a more complete description of the statistics of the field can be obtained with the aid of correlation characteristics. Furthermore, if the field of the antenna is a normal random function, then the description of the statistics with the aid of the correlation functions is complete.

An example of a problem whose solution necessitates knowledge of the correlation functions is the very interesting problem of finding the probability that the directivity pattern of an antenna, in a certain specified sector $\psi_1 - \psi_2$, does not go beyond a definite level, characterized by a function $v(\psi)$ on the interval $\psi_1 - \psi_2$.

Such a problem is of practical interest in estimating the suitability of an antenna on the basis of the side lobes of its pattern.*

*It must be noted that when account is taken of the statistical nature of the field of the antenna, the requirements concerning the side lobes must be formulated approximately as follows: the antenna is assumed suitable if the probability that the directivity pattern of the antenna in the given sector does not go beyond a certain level is not smaller than a specified value.

A solution of this problem and of similar problems is tantamount to determining the functionals of the distribution of real random functions. For a normally distributed random function the approximate value of the functional of the distribution can be determined by determining from the known correlation function the interval of the correlation and by further establishing the probability of joint realization of a series of independent events.

Thus, for normally distributed functions the problem reduces to finding the correlation functions. Unfortunately, the modulus of the complex factor of the system, characterizing the directivity apttern, is not a normally distributed function. In spite of this, the functional of the distribution for the modulus of the multiplier of the system can be obtained, as we shall show below, from the known correlation function for the field.

In the present paper we determine the correlation function of the field of a linear in-phase antenna in the presence of random phase errors of the sources. The correlation function is then used to determine the functional of the distribution of the envelope of the directivity of the pattern. The results obtained in Sections 1 — 3 are equally suitable for the time statistics and for the statistics relative to a family of like antennas. In Section 4 we give a quantitative estimate of the functional for the case when the phase errors of the sources result from inaccuracy in the surface finish of the antenna.

The phase errors are assumed small.

1. INITIAL RELATIONSHIPS

The system multiplier $f(\psi)$ for a linear in-phase antenna with amplitude distribution $A(x)$ has the following form, if account is taken of the random phase errors characterized by the random function $\varphi(x)$: *

$$f(\psi) = \int_{-1}^{+1} A(x)\, e^{j[\varphi(x)+\psi x]}dx,$$

where $\psi = \frac{\pi L}{\lambda} \sin\theta$; θ — angle measured from the normal to the system; λ — length of the wave; L — length of the system.

Assuming the phase errors to be small, we obtain

$$f(\psi) = \int_{-1}^{+1} A(x)\, e^{j\psi x}dx + j\int_{-1}^{+1} \varphi(x)\, A(x)\, e^{j\psi x}dx. \tag{1}$$

Assuming $\overline{\varphi(x)} = 0$, we obtain the average value of the system multiplier

$$\overline{f(\psi)} = \int_{-1}^{+1} A(x)\, e^{j\psi x}dx = f_0(\psi),$$

where $f_0(\psi)$ is the system multiplier in the absence of errors. The field fluctuation is

$$\Delta f(\psi) = j\int_{-1}^{+1} \varphi(x)\, A(x)\, e^{j\psi x}dx. \tag{2}$$

The correlation function of the random complex function $f(\psi)$ can be written, by definition [5], in the form

$$K_f(\psi;\psi_1) = \overline{\Delta f(\psi)\,\Delta f^*(\psi_1)}. \tag{3}$$

The normalized correlation function will be

---

*We have left out the factor L/2, which is of no significance.

$$R_f(\psi;\psi_1) = \frac{K_f(\psi;\psi_1)}{\sqrt{K_f(\psi;\psi)\,K_f(\psi_1;\psi_1)}}. \tag{4}$$

Using (2), we obtain

$$K_f(\psi;\psi_1) = \int_{-1}^{+1}\int \overline{\varphi(x)\,\varphi(x_1)}\,A(x)\,A(x_1)\,e^{j\psi x - j\psi_1 x_1}dx\,dx_1 =$$

$$= \int_{-1}^{+1}\int r\sigma(x)\,\sigma(x_1)\,A(x)\,A(x_1)\,e^{j\psi x - j\psi_1 x_1}dx\,dx_1,$$

where r — correlation function between the values of the phase at the point x and x_1; $\sigma(x)$, mean squared value of the phase $[\sigma^2(x) = \overline{\varphi^2(x)}]$.

We assume further for the correlation function r a form $r = e^{-\frac{(x-x_1)^2}{c^2}}$, where $c = 2\rho/L$ (ρ — correlation interval).

Then

$$K_f(\psi;\psi_1) = \int_{-1}^{+1}\int \sigma(x)\,\sigma(x_1)\,A(x)\,A(x_1)\,e^{-\frac{(x-x_1)^2}{c^2}+j\psi x - j\psi_1 x_1}dx\,dx_1.$$

We confine ourselves further to an examination of the case of uniform amplitude distribution $A(x) = 1$ and error distribution $\sigma(x) = \sigma_0$. In this case

$$K_f(\psi;\psi_1) = \sigma_0^2 \int_{-1}^{+1}\int e^{-\frac{(x-x_1)^2}{c^2}+j\psi x - j\psi_1 x_1}dx\,dx_1. \tag{5}$$

The double integral I in (5) can be calculated exactly. As shown in Appendix 1, the final expression is

$$I(\psi;\psi_1) = \int_{-1}^{+1}\int e^{-\frac{(x-x_1)^2}{c^2}+j\psi x - j\psi_1 x_1}dx\,dx_1 = c\sqrt{\pi}\,\frac{\sin(\psi-\psi_1)}{\psi-\psi_1}\times$$

$$\times\left\{e^{-\frac{\psi_1^2c^2}{4}}\operatorname{Re}\Phi\left(\frac{2}{c}-j\frac{\psi_1 c}{2}\right)+e^{-\frac{\psi^2c^2}{4}}\operatorname{Re}\Phi\left(\frac{2}{c}-j\frac{\psi c}{2}\right)\right\}-2c\,\frac{\cos(\psi-\psi_1)}{\psi-\psi_1}\times$$

$$\times\left\{\left[F\left(\frac{\psi c}{2}\right)-F\left(\frac{\psi_1 c}{2}\right)\right]+\frac{\sqrt{\pi}}{2}\left[e^{-\frac{\psi^2c^2}{4}}\operatorname{Im}\Phi\left(\frac{2}{c}-j\frac{\psi c}{2}\right)-\right.\right.$$

$$\left.\left.-e^{-\frac{\psi_1^2c^2}{4}}\operatorname{Im}\Phi\left(\frac{2}{c}-j\frac{\psi_1 c}{2}\right)\right]\right\}, \tag{6}$$

where

$$\Phi(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}dt; \quad F(z) = e^{-z^2}\int_0^z e^{t^2}dt.$$

Putting $\psi = \psi_1$ in (6), we obtain after simple manipulations

$$I_0(\psi) = \int_{-1}^{+1}\int e^{-\frac{(x-x_1)^2}{c^2}+j\psi(x-x_1)}dx\,dx_1 = \sqrt{\pi}\,c^2 e^{-\frac{\psi^2c^2}{4}}\operatorname{Re}\left[\left(\frac{2}{c}+j\frac{\psi c}{2}\right)\times\right.$$

$$\times \Phi\left(\frac{2}{c} + j\,\frac{\psi c}{2}\right)\Big] - c^2\Big[1 - \psi c F\left(\frac{\psi c}{2}\right) - e^{-\frac{4}{c^2}}\cos 2\psi\Big]. \tag{7}$$

Relations (6) and (7) enable us to calculate the functions

$$K_f = \sigma_0^2 I(\psi; \psi_1), \quad R_f = \frac{I(\psi; \psi_1)}{\sqrt{I_0(\psi)\, I_0(\psi_1)}}$$

for any value of the parameter c, i.e., for any ratio of the correlation interval of the phase errors ρ to the dimension of the antenna L.

It is of interest to consider the character of variation of the normalized correlation function R_f with varying parameter c. For this purpose we put $\psi_1 = 0$ and calculate the quantity $R_f(0, \psi)$, which is equal to

$$R_f(0, \psi) = \frac{I(0, \psi)}{\sqrt{I_0(0)\, I_0(\psi)}}, \tag{8}$$

where

$$I(0, \psi) = c\sqrt{\pi}\,\frac{\sin\psi}{\psi}\Big[\Phi\left(\frac{2}{c}\right) + e^{-\frac{\psi^2 c^2}{4}}\,\mathrm{Re}\,\Phi\left(\frac{2}{c} - j\,\frac{\psi c}{2}\right)\Big] -$$

$$- 2c\,\frac{\cos\psi}{\psi}\Big[F\left(\frac{\psi c}{2}\right) + \frac{\sqrt{\pi}}{2}\,e^{-\frac{\psi^2 c^2}{4}}\,\mathrm{Im}\,\Phi\left(\frac{2}{c} - j\,\frac{\psi c}{2}\right)\Big]; \tag{9}$$

$$I_0(0) = 2\sqrt{\pi}\, c\Phi\left(\frac{2}{c}\right) - c^2\left(1 - e^{-\frac{4}{c^2}}\right), \tag{10}$$

$I_0(\psi)$ is determined from (7).

Expressions (7) and (9) contain the probability integral of complex argument. It is more convenient to use in the calculations, in lieu of $\Phi(z)$, the function $W(z)$, defined by

$$W(z) = e^{-z^2}\left(1 + \frac{2j}{\sqrt{\pi}}\int_0^z e^{t^2} dt\right),$$

for which there are detailed tables [6]. To use these tables it is necessary to replace the function $\Phi(z)$ in (7) and (9) by the function $W(z)$. Taking into account the relation $\Phi(z) = e^{-z^2} W(-jz) - 1$, it is easy to transform (7) and (9) into

$$I_0(\psi) = 2c\sqrt{\pi}\, e^{-\frac{\psi^2 c^2}{4}} + c^2\sqrt{\pi}\, e^{-\frac{4}{c^2}}\,[(yV - xU)\sin 2\psi - (yU + xV)\cos 2\psi] -$$

$$- c^2\Big[1 - \psi c F\left(\frac{\psi c}{2}\right) - e^{-\frac{4}{c^2}}\cos 2\psi\Big], \tag{7a}$$

$$I(0, \psi) = c\sqrt{\pi}\,\frac{\sin\psi}{\psi}\left\{\Phi\left(\frac{2}{c}\right) + e^{-\frac{\psi^2 c^2}{4}} - e^{-\frac{4}{c^2}}\,[U\cos 2\psi - V\sin 2\psi]\right\} -$$

$$- 2c\,\frac{\cos\psi}{\psi}\left\{F\left(\frac{\psi c}{2}\right) - \frac{\sqrt{\pi}}{2}\,e^{-\frac{4}{c^2}}\,[V\cos 2\psi + U\sin 2\psi]\right\}, \tag{9a}$$

where

$$U = \mathrm{Re}\, W(z); V = \mathrm{Im}\, W(z); z = x + jy = \frac{\psi c}{2} + j\,\frac{2}{c}.$$

When $|z| \gg 1$ and $|\arg z| < \frac{\pi}{2} - \delta (\delta > 0)$, using the asymptotic expression for the probability integral

$$\Phi(z) \simeq 1 - \frac{e^{-z^2}}{\sqrt{\pi}\, z} - \frac{e^{-z^2}}{2\sqrt{\pi}\, z^3},$$

we obtain from (8) and (10) the following expressions

$$I_0(\psi) \simeq 2\sqrt{\pi}\, ce^{-\frac{\psi^2c^2}{4}} - c^2\left[1 - \psi cF\left(\frac{\psi c}{2}\right)\right] + \\ + \frac{c^2}{2} e^{-\frac{4}{c^2}} \frac{\left[\left(\frac{4}{c^2} - \frac{\psi^2c^2}{4}\right)\cos 2\psi - 2\psi \sin 2\psi\right]}{\left[\left(\frac{4}{c^2} - \frac{\psi^2c^2}{4}\right)^2 + 4\psi^2\right]}, \tag{7b}$$

$$I(0, \psi) \simeq c\sqrt{\pi}\frac{\sin\psi}{\psi}\left[\Phi\left(\frac{2}{c}\right) - e^{-\frac{\psi^2c^2}{4}}\right] - 2c\frac{\cos\psi}{\psi}F\left(\frac{\psi c}{2}\right) + \\ + \frac{ce^{-\frac{4}{c^2}}}{\frac{4}{c^2} + \frac{\psi^2c^2}{4}}\left[\frac{2}{c}\frac{\sin\psi}{\psi} + \frac{c}{2}\cos\psi\right]. \tag{9b}$$

Using the exact expressions (10) and (7a), (9a) or (when $|z| \gg 1$) the asymptotic expressions (7b) and (9b), we can calculate with the aid of formula (8) the value of $R_f(0, \psi)$ for different values of the parameter c.

The result of the calculation of $R_f(0, \psi)$ for several values of the parameter c are illustrated in Fig. 1. As can be seen from the figure, even when $c \sim 3$ ($\rho \sim \frac{3}{2}L$) the normalized correlation function $R_f(0, \psi)$ has a modulus close to unity in the entire interval of variation of ψ ($\psi \leq 10$) which we are considering. As c decreases, the correlation distance also decreases.

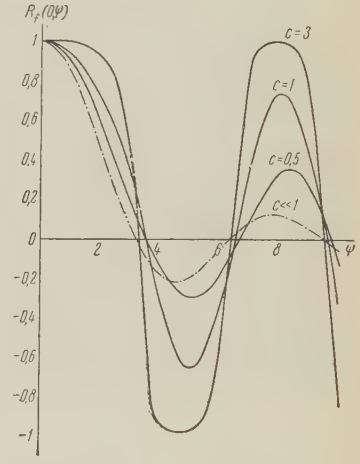


Fig. 1.

In the limiting case $c \ll 1$, by retaining in the general relations (6) and (7) only the first term of the asymptotic expansion in c, we obtain

$$K_f(\psi; \psi_1) \simeq 2\sqrt{\pi}\, c\, \sigma_0^2 \frac{\sin(\psi - \psi_1)}{\psi - \psi_1},$$

$$R_f(\psi; \psi_1) \simeq \frac{\sin(\psi - \psi_1)}{\psi - \psi_1}. \tag{11}$$

It is seen from (11) that when $c \ll 1$ the correlation function depends only on the difference of the argument $\psi - \psi_1$, and the correlation interval for the function $f(\psi)$ is a quantity of the order of π.

The function $\sin\psi/\psi$, which has a value $R_f(0, \psi)$ when $c \ll 1$, is shown dashed in Fig. 1.

The case $c \ll 1$ is of particular interest, for it arises in the study of antenna family statistics involving manufacturing inaccuracies in antenna surface finish.

The correlation interval ρ of the phase errors, connected with the inaccuracy of the manufacture of the antenna, is as a rule of the order of λ [3], and the dimensions of the antenna L are much greater than λ.

The condition $c \ll 1$ is usually satisfied in the case when the phase fluctuations of the sources are the consequence of time-varying deformations of the antenna surface.

2. FUNCTIONAL OF THE DISTRIBUTION OF THE DIRECTIVITY PATTERN

In the study of the function of the distribution of the directivity pattern we confine ourselves to the case $c \ll 1$ ($\rho \ll L$). If $c \ll 1$, then the second integral in (1) can be represented in the form of a sum of a large number of weakly dependent terms. From this, on the

basis of the central limit theorem, we can conclude that the random function $f(\psi)$ has a normal distribution law.*

The latter, in particular, signifies that:

a) for any fixed ψ the real and imaginary parts of the complex multiplier $f(\psi)$ obey a two-dimensional normal distribution;

b) the vanishing of the correlation function at a certain value $\psi - \psi_1$ signifies the independence of the random quantities $f(\psi)$ and $f(\psi_1)$.

As noted earlier, when $c \ll 1$ the correlation interval for the function $f(\psi)$ has a value on the order of π. It follows therefore that the values of the modulus (or phase) of the function $f(\psi)$ at points separated by intervals on the order of π will also be independent.

We note further that when $A(x) = 1$ the directivity pattern of the system, in the absence of errors, will be

$$f_0(\psi) = 2\frac{\sin\psi}{\psi}.$$

The first maximum corresponds to $\psi = 0$. The position of the remaining maxima of the modulus $f_0(\psi)$ is determined by the roots of the transcendental equation $\tan\psi = \psi$. The approximate roots of this equation are

$$\psi_k \simeq (2k+1)\frac{\pi}{2}, \quad k = 1, 2, \ldots.$$

Thus, the maxima of the function $|f_0(\psi)|$ are divided by intervals on the order of π.

Going over to an examination of the functional of the distribution of the directivity pattern in the presence of error, we note that in the study of the problem of the side lobes we are interested essentially in the envelope of the directivity pattern. In the case of small errors the behavior of the envelope is determined by the values of the modulus of $f(\psi)$ at the points ψ_k. It is clear therefore that in order to find the approximate value of the functional of the distribution of the directivity pattern (more accurately, the functional of the distribution of the envelope) it is first necessary to calculate the probability P_k that the values $|f(\psi)|$ at the points ψ_k do not exceed the values of the function $v(\psi)$ at the same points, and then taking into account the independence of the values of the modulus at the points ψ_k, we take the product of the resultant values of P_k.

Thus, it is necessary first of all to find the integral law of distribution of the envelope (modulus) at the points ψ_k. The random quantity $|f(\psi)|$ is the length of a vector with components $a = \mathrm{Re}\, f(\psi)$ and $b = \mathrm{Im}\, f(\psi)$. As follows from (1), the values of a and b are determined from the expressions**

$$a = \int_{-1}^{+1} A(x)\cos\psi x\,dx - \int_{-1}^{+1} A(x)\varphi(x)\sin\psi x\,dx,$$

$$b = \int_{-1}^{+1} A(x)\sin\psi x\,dx + \int_{-1}^{+1} A(x)\varphi(x)\cos\psi\, x\,dx = \int_{-1}^{+1} A(x)\varphi(x)\cos\psi x\,dx.$$

The values of a and b have a normal distribution with mean values

$$\bar{a} = \int_{-1}^{+1} A(x)\cos\psi x\,dx, \quad \bar{b} = 0$$

and with dispersions (see Appendix II)

---

*In order for the function $f(\psi)$ to be a normally distributed random function for arbitrary c, it is necessary that the random function $\varphi(x)$, which characterizes the phases errors, be normal.

**The function A(x) is assumed even.

$$\sigma_a^2 \simeq \sqrt{\pi}\, ce^{-\frac{\psi^2 c^2}{4}} \int_{-1}^{+1} A^2(x)\, \sigma^2(x) \sin^2 \psi x\, dx,$$
$$\sigma_b^2 \simeq \sqrt{\pi}\, ce^{-\frac{\psi^2 c^2}{4}} \int_{-1}^{+1} A^2(x)\, \sigma^2(x) \cos^2 \psi x\, dx. \tag{12}$$

The correlation moment of the quantities a and b will be

$$K_{ab} \simeq \sqrt{\pi}\, ce^{-\frac{\psi^2 c^2}{4}} \int_{-1}^{+1} A^2(x)\, \sigma^2(x) \sin \psi x \cos \psi x\, dx.$$

For even variations of A(x) and $\sigma(x)$, the correlation moment vanishes. The vanishing of the correlation moment for a normal distribution law of the aggregate of the values a and b means that they are independent.

For the case A(x) = 1 and $\sigma(x)$ relations (12) can be rewritten as

$$\sigma_a^2 \simeq \sqrt{\pi}\, ce^{-\frac{\psi^2 c^2}{4}} \sigma_0^2 \left(1 - \frac{\sin 2\psi}{2\psi}\right), \quad \sigma_b^2 \simeq \sqrt{\pi}\, ce^{-\frac{\psi^2 c^2}{4}} \sigma_0^2 \left(1 + \frac{\sin 2\psi}{2\psi}\right).$$

At the points $\psi_k \simeq (2k+1)\frac{\pi}{2}$ we have $\sin 2\psi \simeq 0$, i.e., the dispersions of σ_a and σ_b at these points, namely σ_{ak} and σ_{bk}, are the same:

$$\sigma_{ak}^2 \simeq \sigma_{bk}^2 \simeq \sqrt{\pi}\, c\sigma_0^2 e^{-\frac{\psi_k^2 c^2}{4}} = \sigma_k^2. \tag{13}$$

Thus, the question of finding the law of distribution of the modulus $f(\psi)$ at the points ψ_k reduces to the problem of the distribution of the length of a vector whose components a and b are independent and have a normal distribution with parameters $(\bar{a}, \sigma_k)$ and $(0, \sigma_k)$.

The solution to this problem is well known [7]. The distribution function of the length of a vector is a generalized Rayleigh distribution function

$$\omega_k(\rho) = \frac{\rho}{\sigma_k^2} e^{-\frac{\rho^2 + \bar{a}_k^2}{2\sigma_k^2}} I_0\left(\frac{\bar{a}_k \rho}{\sigma_k^2}\right), \tag{14}$$

where

$$\rho = |f(\psi)|; \quad \bar{a}_k = 2\frac{\sin \psi_k}{\psi_k} \simeq \frac{4}{(2k+1)\pi};$$

I_0 — modified Bessel function of zero order.

The integral distribution law $|f(\psi)|$ at the points ψ_k will be

$$P_k(|f(\psi_k)| \leqslant v(\psi_k)) = e^{-\frac{[v(\psi_k)]^2 + \bar{a}_k^2}{2\sigma_k^2}} \sum_{n=1}^{\infty} \left(\frac{v(\psi_k)}{\bar{a}_k}\right)^n I_n\left(\frac{\bar{a}_k v(\psi_k)}{\sigma_k^2}\right). \tag{15}$$

The curves of the integral distribution law are shown in Fig. 2.

For the case when $v(\psi)$ coincides with the theoretical envelope, we have $v(\psi_k) = \bar{a}_k$ and the values of P_k are determined from

$$P_k = e^{-\frac{\bar{a}_k^2}{\sigma_k^2}} \sum_{n=1}^{\infty} I_n\left(\frac{\bar{a}_k^2}{\sigma_k^2}\right).$$

If the errors are small, so that the inequality $\bar{a}_k/\sigma_k \gg 1$ holds, then the generalized Rayleigh law (14) goes over into the normal law with parameters a_k and σ_k. The integral distribution function will in this case be

$$P_k \simeq \frac{1}{\sqrt{2\pi}\,\sigma_k} \int_{-\infty}^{v(\psi_k)} e^{-\frac{(\xi-\bar{a}_k)^2}{2\sigma_k^2}} d\xi = F\left(\frac{v(\psi_k) - \bar{a}_k}{\sigma_k}\right). \tag{16}$$

The curve of the integral normal distribution law F(x) is shown in Fig. 3.

The functional of the distribution of the envelope is determined as the product of the values of P_k. It is necessary to bear in mind here that for each k we have two points, situated symmetrically on both sides of the direction of the maximum ($\psi = 0$). In the general case the values of $v(\psi_k)$ at these points may not be equal. Consequently the values of P_k at these points will also be different.

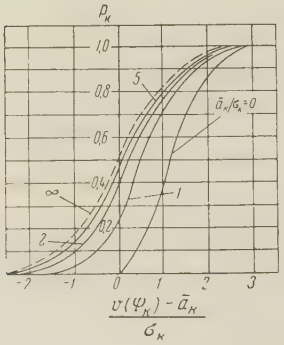


Fig. 2

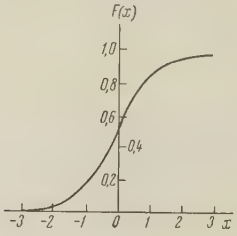


Fig. 3

However, usually the function $v(\psi)$ is even. In this case the functional of the distribution of the envelope can be written in the form

$$P_{v(\psi)} = \prod_{k=1}^{N} P_k^2. \tag{17}$$

3. QUANTITATIVE ESTIMATE OF P

Going over to an analysis of certain examples, it is necessary to note first that the aperture method of calculating antennas is used usually to find a principal lobe and some of the side lobes that are closest to it. This circumstance predetermines the interval of values of ψ, for which it is advantageous to use the result obtained above. If the errors are in this case so small that for the extreme investigated point ψ_N the inequality $\bar{a}_N/\sigma_N \gg 1$ is satisfied, we determine P from the formula

$$P_{v(\psi)} = \prod_{k=1}^{N} F_k^2, \tag{18}$$

where F_k is determined by the relation (16), i.e.,

$$F_k = F(\delta_k), \tag{19}$$

where

$$\delta_k = \frac{v(\psi_k) - \bar{a}_k}{\sigma_k}.$$

In particular, the probability that on the section $\pm\psi_N$ the diagram will not go outside the theoretical envelope ($\delta_k = 0$), will be

$$P = \prod_{k=1}^{N} F^2(0) = 0{,}25^N, \tag{20}$$

where N is the number of side lobes considered.

Let us further consider by way of a specific example the case when the field fluctuations are the consequence of inaccuracy in the surface finish of a mirror antenna. In this case we can use for estimates of σ_0 the argument advanced in reference [3]. Assuming the relative tolerance in the surface of the antenna to be specified (in accordance with technological capabilities), the author of that paper obtains the following expression for the mean square deviation of the phases of the sources:

$$\sigma_0 \simeq 5 \frac{L}{\lambda} 10^{-m}, \tag{21}$$

where L — linear dimension of the antenna; m — a quantity characterizing the relative tolerance in the manufacture of the antenna. It is determined by the relation

$$10^{-m} = \frac{\varepsilon}{L},$$

where ϵ — tolerance in the surface finish of the antenna, which can be guaranteed with a probability 0.99.

Let us now determine the quantity c. As was noted above, the correlation interval usually is on the order of λ. Since the total length of the antenna in dimensionless units is 2 (x varies from −1 to +1), we have

$$c \simeq 2 \frac{\lambda}{L}. \tag{22}$$

Using (13), (21), and (22) we obtain

$$\sigma_k \simeq 9{,}4 \cdot 10^{1-m} \sqrt{\frac{L}{\lambda}} e^{-\frac{1}{2}\left[\frac{\lambda}{L}\frac{(2k+1)\pi}{2}\right]^2} \tag{23}$$

We confine ourselves to $k \leq 3$ and put $L/\lambda > 10$. Then

$$\begin{aligned} e^{-\frac{1}{2}\left[\frac{\lambda}{L}\frac{(2k+1)\pi}{2}\right]^2} &\simeq 1, \\ \sigma_k &\simeq 10^{-m}\sqrt{\frac{L}{\lambda}}. \end{aligned} \tag{24}$$

Assuming, for example, furthermore than m = 3* and recognizing that $\bar{a}_k = 4/(2k+1)\pi$, we find that in the range

$$10 < \frac{L}{\lambda} < 50 \tag{25}$$

we have the inequality $\bar{a}_k/\sigma_k \gg 1$. Thus, when $k \leq 3$, m = 3 and L/λ satisfies the inequality (25), the value of P can be calculated from formulas (18) and (19).

Fig. 4 shows three functions $v(\psi)$. The first, $v_1(\psi)$, coincides with the theoretical envelope. the probability that the directivity pattern will not go beyond $v_1(\psi)$ in the portion $\pm\psi_3$.

---

*The value m = 3 corresponds to a very good antenna surface finish.

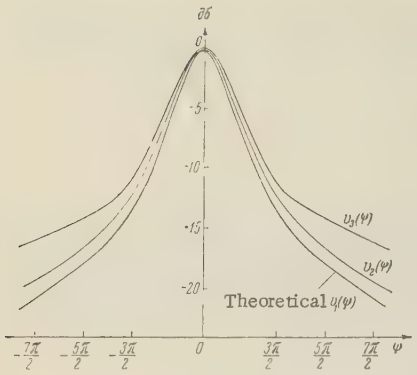


Fig. 4

It will, in accordance with (20), have a value $P_{v1(\psi)} \simeq 0.016 = 2\%$.

The curves $v_{2,3}(\psi)$ are drawn through the points $\bar{a}_k + 2\sigma_k$. With this $v_2(\psi)$ corresponds to $L/\lambda = 10$; $v_3(\psi)$ corresponds to $L/\lambda = 50$. In both cases we have $\delta_k = 2$ and in accordance with (18) and (19) we get $P = F^6(2)$. Using the tables for the function F [7], we obtain $F(2) = 0.977$ and $P = 0.86$.

Thus, we can state that, with a probability of 86 percent, the directivity pattern of the antenna in the portion $\pm\psi_3(\pm\frac{7\pi}{2})$ will not go beyond the limits $v_2(\psi)$ for a system of length 10λ and $v_3(\psi)$ for a system of length 50λ.

The table lists the values of σ_k calculated by formula (23) for $L/\lambda = 10$ and $L/\lambda = 50$. The table also indicates the levels of the side lobes, corresponding to the curves $v_2(\psi)$ and $v_3(\psi)$.

It must be noted that if the inequality $\delta_k\sigma_k/\bar{a}_k \ll 1$ is satisfied at the point ψ_k, then the change in level of the side lobe radiation compared with the theoretical value will be

K	σ_k		Side load level, db		
	$L/\lambda = 10$	$L/\lambda = 50$	Theoretical $v_1(\psi)$	$v_2(\psi)$	$v_3(\psi)$
1	0.027	0.065	-13.3	-12.4	-11.1
2	0.021	0.066	-17.7	-16.6	-14.3
3	0.016	0.066	-20.8	-19.4	-16.1

$$\Delta_k(\partial\delta) = 20\lg\frac{\bar{a}_k + \delta_k\sigma_k}{\bar{a}_k} =$$

$$= 8{,}6\ln\left(1 + \frac{\delta_k\sigma_k}{\bar{a}_k}\right) \simeq 8{,}6\frac{\delta_k\sigma_k}{\bar{a}_k}.$$

From this, in particular, it follows that Δ_k (in decibels) is proportional to the tolerance ϵ. This must be kept in mind in the analysis of the data in the table. The relatively small growth in the level of the side lobes is connected with the rather high accuracy which we have assumed for the system manufactured. The assumption $m = 3$ means, for example, that a system one meter long is prepared with a tolerance $\epsilon = 1$ mm. An increase in the tolerance ϵ by a factor, say, of two will increase Δ_k(db) also by a factor of two.

The example given above illustrates the order of the calculation of the value of P for a specified function $v(\psi)$, for a specified interval of values of ψ, and for a known precision in the manufacture of the antenna m.

The results obtained above can be used also to solve the inverse problem — determine the necessary accuracy of antenna manufacture for specified side lobes.

Let us consider by way of an example the following problem: it is required to determine the accuracy of an antenna of length L, under the condition that the probability that the first side lobe does not exceed $-M$ db is P.

From (17) we get $P = P_1^2$, i.e., $P_1 = \sqrt{P}$.

Let us determine the value of $v(\psi)$:

$$20\lg\frac{v(\psi_1)}{2} = -M.$$

Hence

$$v(\psi_1) = 2\cdot 10^{-\frac{M}{20}}.$$

From Fig. 2 or from the tables we obtain for the function P_1 the value of σ_1. If we expect here to have $\bar{a}_1/\sigma_1 \gg 1$ (which must be verified subsequently), we can use Fig. 3 or the table for the function F.

Using further the relation (23) or (24), we find the value of m and accordingly ϵ. Thus, if we use the expression (24), then

$$10^m \simeq 10 \sqrt{\frac{L}{\lambda}} \frac{1}{\sigma_1}.$$

Let, for example $L/\lambda = 40$, $P = 0.9$, $M = -10$ db. Then $P_1 = 0.95$, $v(\psi_1) = 2 \cdot 10^{-\frac{1}{2}} = 0.635$.

Putting $\bar{a}_1/\sigma_1 \gg 1$*, we get $P_1 = f(\delta_1) = 0.95$. Therefore $\delta_1 = 1.65$ and

$$\sigma_1 = \frac{v(\psi_1) - \bar{a}_1}{\delta_1} = 0{,}127.$$

The necessary accuracy of manufacture will be $10^m \simeq 10 \sqrt{\frac{L}{\lambda}} \frac{1}{\sigma_1} \simeq 500$, $10^{-m} = \epsilon/L \simeq$ $\simeq 0.002$, $m \simeq \lg 500 = 2.7$, $\epsilon \simeq 0.002L = 0.08\lambda$.

In conclusion we note that the results of the present investigation can be used to study the correlation properties of the diffraction image in focusing systems.

APPENDIX 1

Calculation of the Integral I

In the determination of the correlation function $K_f(\psi; \psi_1)$ it is necessary to evaluate an integral of the form

$$I = \int_{-1}^{+1}\int e^{-\frac{(x-x_1)^2}{c^2} + j\psi x - j\psi_1 x_1} dx\, dx_1.$$

Let us change from the variable x, x_1, to new variables u, v:

$$u = \frac{x_1 - x}{\sqrt{2}}, \quad v = \frac{x_1 + x}{\sqrt{2}}.$$

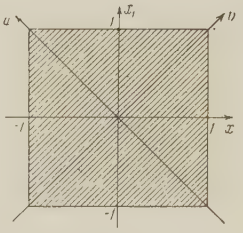


Fig. 5

The change to the variables u and v signifies a rotation of the coordinate system, as shown in Fig. 5. Recognizing that the double integral is taken over the shaded square in Fig. 5, we have

$$I = \int_{-1}^{+1}\int e^{-\frac{(x-x_1)^2}{c^2} + j\psi x - j\psi_1 x_1} dx\, dx_1 =$$

$$= \int_0^{\sqrt{2}} e^{-\frac{2u^2}{c^2} - j\frac{u}{\sqrt{2}}(\psi+\psi_1)} du \int_{-(\sqrt{2}-u)}^{\sqrt{2}-u} e^{j\frac{v}{\sqrt{2}}(\psi-\psi_1)} dv +$$

$$+ \int_{-\sqrt{2}}^{0} e^{-\frac{2u^2}{c^2} - j\frac{v}{\sqrt{2}}(\psi+\psi_1)} du \int_{-(\sqrt{2}+u)}^{\sqrt{2}+u} e^{j\frac{v}{\sqrt{2}}(\psi-\psi_1)} dv =$$

$$= \frac{4\sqrt{2}}{\psi - \psi_1} \int_0^{\sqrt{2}} \cos\frac{u(\psi+\psi_1)}{\sqrt{2}} \sin\frac{(\sqrt{2}-u)(\psi-\psi_1)}{\sqrt{2}} e^{-\frac{2u^2}{c^2}} du.$$

---

*In this example $\bar{a}_1/\sigma_1 \simeq 3.3$ and the function P_1 can be replaced by the function F.

Going to a new variable $\frac{\sqrt{2u}}{c}$, we obtain

$$I = \frac{2c \sin(\psi - \psi_1)}{\psi - \psi_1}\left[\int_0^{2/c} e^{-y^2}\cos c\psi_1 y\, dy + \int_0^{2/c} e^{-y^2}\cos c\psi y\, dy\right] -$$

$$- \frac{2c\cos(\psi - \psi_1)}{\psi - \psi_1}\left[\int_0^{2/c} e^{-y^2}\sin c\psi y\, dy - \int_0^{2/c} e^{-y^2}\sin c\psi_1 y\, dy\right].$$

The integrals in this expression can be readily calculated. Leaving out the intermediate steps, we write the following results

$$\int_0^{2/c} e^{-y^2}\cos c\psi y\, dy = \frac{\sqrt{\pi}}{2} e^{-\frac{\psi^2 c^2}{4}} \operatorname{Re}\Phi\left(\frac{2}{c} - j\frac{\psi c}{2}\right),$$

$$\int_0^{2/c} e^{-y^2}\sin c\psi y\, dy = F\left(\frac{\psi c}{2}\right) + \frac{\sqrt{\pi}}{2} e^{-\frac{\psi^2 c^2}{4}} \operatorname{Im}\Phi\left(\frac{2}{c} - j\frac{\psi c}{2}\right),$$

where

$$\Phi(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2} dt; \qquad F(z) = e^{-z^2}\int_0^z e^{t^2} dt.$$

Using these results we obtain as the final expression for the integral I:

$$I = \int_{-1}^{+1}\int e^{-\frac{(x-x_1)^2}{c^2} + j\psi x - j\psi_1 x_1} dx\, dx_1 = c\sqrt{\pi}\frac{\sin(\psi - \psi_1)}{\psi - \psi_1}\left\{e^{-\frac{\psi_1^2 c^2}{4}} \operatorname{Re}\Phi\left(\frac{2}{c} - j\frac{\psi_1 c}{2}\right) + \right.$$

$$\left. + e^{-\frac{\psi^2 c^2}{4}} \operatorname{Re}\Phi\left(\frac{2}{c} - j\frac{\psi c}{2}\right)\right\} - 2c\frac{\cos(\psi - \psi_1)}{\psi - \psi_1}\left\{\left[F\left(\frac{\psi c}{2}\right) - F\left(\frac{\psi_1 c}{2}\right)\right] + \right.$$

$$\left. + \frac{\sqrt{\pi}}{2}\left[e^{-\frac{\psi^2 c^2}{4}} \operatorname{Im}\Phi\left(\frac{2}{c} - j\frac{\psi c}{2}\right) - e^{-\frac{\psi_1^2 c^2}{4}} \operatorname{Im}\Phi\left(\frac{2}{c} - j\frac{\psi_1 c}{2}\right)\right]\right\}.$$

APPENDIX 2

Calculation of the Dispersion of the Components of the System Multiplier $f(\psi)$

We confine ourselves to the calculation of the dispersion of the random quantity

$$a = \operatorname{Re} f(\psi) = \bar{a} - \int_{-1}^{+1} A(x)\,\varphi(x)\sin\psi x\, dx,$$

$$\sigma_a^2 = \int_{-1}^{+1}\int A(x)\,A(x_1)\,\overline{\varphi(x)\,\varphi(x_1)}\sin\psi x\sin\psi x_1\, dx\, dx_1 =$$

$$= \int_{-1}^{+1}\int A(x)\,A(x_1)\,r\sigma(x)\,\sigma(x_1)\sin\psi x\sin\psi x_1\, dx\, dx_1 =$$

$$= \int_{-1}^{+1} A(x)\,\sigma(x)\sin\psi x\, dx \int_{-1}^{+1} A(x_1)\,\sigma(x_1)\sin\psi x_1 e^{-\frac{(x-x_1)^2}{c^2}} dx_1.$$

If the correlation interval ρ is much smaller than the linear dimensions of the system L, i.e., if $c \ll 1$, then the region of significance in the evaluation of the internal integral lies

near $x = x_1$, and therefore the limits in the internal integral can be extended to $\pm\infty$.

Assume further that $A(x)\sigma(x)$ is a slowly varying function, we have

$$\sigma_a^2 = \int_{-1}^{+1} A(x)\,\sigma(x)\sin\psi x\,dx \int_{-1}^{+1} A(x_1)\,\sigma(x_1)\sin\psi x_1 e^{-\frac{\gamma(x-x_1)^2}{c^2}}\,dx_1 \simeq$$
$$\simeq \int_{-1}^{+1} A^2(x)\,\sigma^2(x)\sin\psi x\,dx \int_{-\infty}^{+\infty} \sin\psi x_1 e^{-\frac{(x-x_1)^2}{c^2}}\,dx_1.$$

Using the formula [8]

$$\int_{-\infty}^{\infty} e^{-q^2x^2} \begin{matrix}\sin\\ \cos\end{matrix} [p(x+\lambda)]\,dx = \frac{\sqrt{\pi}}{q} e^{-\frac{p^2}{4q^2}} \begin{matrix}\sin\\ \cos\end{matrix} p\lambda,$$

we obtain finally

$$\sigma_a^2 \simeq \sqrt{\pi}\,ce^{-\frac{\psi^2c^2}{4}} \int_{-1}^{+1} A^2(x)\,\sigma^2(x)\sin^2\psi x\,dx.$$

Analogously we calculate σ_b^2 and the correlation moment of the quantities a and b.

REFERENCES

1. J. Ruze. Effect of aperture errors on antenna radiation patterns. Voprosy radiolokatsionnoy tekhniki, 1956, 2, 98.
2. Ye.N. Gil'bert, S.P. Morgan. Calculation of optimal antenna grid in the presence of random changes. Ibid, 1955, 6, 55.
3. J. Robius. Influence de la precision de fabrication d'une antenne sur ses performances. Ann. radioelectr., 1956, 11, 43, 26.
4. V.I. Talanov, N.M. Sheronova. On the influence of random errors in the distribution of sources on the directivity pattern of traveling wave antennas. Izv. vuzov MVO SSSR (Radiofizika), 1959, 2, 3, 424.
5. V.S. Pugachev. Theory of random functions and its application to automatic ontrol problems, GITTL, 1957.
6. V.M. Fadeeva, N.M. Terent'ev. Table of values of the probability integral for complex arguments, GITTL, 1954.
7. B.R. Levin. Theory of random processes and its use in radio. Izd. Sovetskoye radio, 1957.
8. I.M. Ryzhik. Tables of integrals, sums, series, and products. GTI, 1948.

Received by editor 10 May 1961

DIFFRACTION OF WAVES ON A SLOT AND ON A RIBBON ORIENTED PERPENDICULAR TO AN IMPEDANCE PLANE

M.D. Khaskind

A method is developed for constructing exact solutions of two-dimensional problems on the diffraction of E modes on a slot and on ideally conducting ribbon, oriented perpendicular to an impedance plane. The general method is based on the analysis of the properties of the special functional combination introduced in [1, 2]. Particular attention is paid to the limiting solutions, which are represented in an effective form. The results obtained are compared with the approximate solutions discussed in [3].

1. INITIAL RELATIONSHIPS

The electromagnetic field above an impedance plane $z = 0$ is investigated with the aid of the magnetic vector $\Pi_m = \varphi(y, z)\, x^0$ where x^0 is a unit vector of the x axis and the time factor $\exp i\omega t$ is left out. The intensities of the electric and magnetic fields are represented by the following relations, in the practical rationalized system of units:

$$\mathbf{E} = -ik\rho_0 (\operatorname{grad} \varphi \times x^0), \quad E_x = 0, \quad E_y = -ik\rho_0 \frac{\partial \varphi}{\partial z}, \quad E_z = ik\rho_0 \frac{\partial \varphi}{\partial y}, \tag{1}$$

$$H_x = k^2 \varphi, \quad H_y = H_z = 0 \quad \left(\rho_0 = \left(\frac{\mu_0}{\varepsilon_0}\right)^{1/2} = 120\,\pi\ \text{ом}\right),$$

where $k = \omega(\varepsilon_0 \mu_0)^{1/2}$ — wave number in free space.

On the impedance plane we have the condition

$$E_y = i\rho_0 Z H_x \quad \text{for} \quad z = 0 \ (Z > 0),$$

which for a scalar function φ assumes the form

$$\frac{\partial \varphi}{\partial z} + p\varphi = 0 \quad \text{for} \quad z = 0 \quad (p = kZ). \tag{2}$$

The analysis that follows is based on an examination of the properties of the functional combination

$$\frac{\partial \varphi}{\partial z} - p\varphi = \frac{\partial f}{\partial z}, \tag{3}$$

where the function $f(y, z)$, like the function $\varphi(y, z)$ satisfies the wave equation and, in accordance with (2), the condition

$$\frac{\partial f}{\partial z} = 0 \quad \text{for} \quad z = 0, \tag{4}$$

which enables us to continue the function f into the lower half-space in even fashion;

consequently the function f is regular and single-valued in all of the space outside the cylindrical surface bounded by the transverse contour L = L*, where L is the transverse contour of the cylindrical surface enclosing the sources, and L* is the mirror image of the contour L in the lower half plane.

If f(y, z) is a definite function, then, as shown in [1, 2], the reverse transition to the function $\varphi(y, z)$, with account of the asymptotic behavior, can be carried out in unique fashion. We use below the following formula for the transition

$$\varphi = e^{-pz} \int_{-\infty}^{z} e^{p\zeta} \frac{\partial f}{\partial \zeta} d\zeta + i \frac{p}{h} A(\pm h, p)\, e^{-pz \mp ihy}, \tag{5}$$

$$A(\pm h, p) = \frac{1}{2} \int_{L+L^*} e^{pz \pm ihy} \left[\frac{\partial f}{\partial n} - f(\pm ih \cos(n, y) + p \cos(n, z)) \right] dl$$

$$(h^2 = k^2 + p^2), \tag{6}$$

where the straight-line path of integration $(-\infty, z)$ in (5) does not cross the contour L = L*. The upper sign is taken for points situated to the right of this contour, and the lower one for points on its left. The second term in (5) determines here the excited field of surface waves, while the function $a(\pm h, p)$ represents the complex amplitude of these waves.

The mean electromagnetic energy carried by the surface waves through a unit transverse segment at $y \to \pm \infty$ is determined in simple form [3]

$$P_{\pm} = \rho_0 \frac{pk^3}{4h} |A(\pm h, p)|^2. \tag{7}$$

The total complex power is determined from the general formula

$$P_k = \int_L E_l H_x^* dl = i\rho_0 k^3 \int_L \varphi^* \frac{\partial \varphi}{\partial n} dl, \tag{8}$$

where the tangential component $E_l = i\rho_0 k \partial\varphi/\partial n$ of the electric field intensity vanishes on the perfectly conducting parts of the contour L; on the open parts this component is a specified function or else must be determined when 'joining' the internal and external fields.

2. DIFFRACTION OF WAVES ON A SLOT

Let us consider a rectangular slot between a perfectly conducting plane and impedance plane perpendicular to it (Fig. 1), and let us assume that an arbitrary incident field is specified on the side $y < 0$. This field, with account of reflection from the impedance plane, is characterized by a function $\varphi_0(y, z)$. In particular, for incident plane waves we have

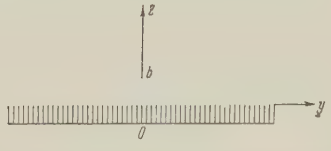


Fig. 1.

$$\varphi_0 = e^{-ik(y \cos\beta - z \sin\beta)} + Re^{-ik(y \cos\beta + z \sin\beta)},$$
$$R = \frac{ik \sin\beta + p}{ik \sin\beta - p}, \tag{9}$$

where β is the elevation angle and R the coefficient of reflection from the impedance plane. If $\beta = i\alpha_0$, i.e., if $k \operatorname{ch}\alpha_0 = h$, $k \operatorname{sh}\alpha_0 = p$, and R = 0, then the incident wave turns into a surface wave with a function $\varphi_0(y,z)$ in the form

$$\varphi_0 = e^{-ihy - pz}. \tag{10}$$

Let the function $\varphi_+ = \varphi(y, z)$ $(y > 0)$ define the transmitted waves; it is then easy to see that when $y < 0$ the electric magnetic field can be defined with the aid of the sum $\varphi_- = \varphi_0(y, z) + \varphi_0(-y, z) - \varphi(-y, z)$, where the scalar function $\varphi_0(-y, z)$ corresponds to

waves reflected from the perfectly conducting plane y = 0. The continuity of the entire solution and the condition $E_z = 0$ on a perfectly conducting plane lead to the relations

$$\varphi(0, z) = \varphi_0(0, z) \quad \text{for} \quad 0 < z < b; \qquad \frac{\partial \varphi}{\partial y} = 0 \quad \text{for} \quad y = 0,\ b < z < \infty. \tag{11}$$

These conditions enable us to continue the function from the region y > 0 to the region y < 0 in even fashion, and consequently, the values of $\partial\varphi/\partial y$ have different signs on opposite sides of the segment (p, b) of the z axis.

Let us consider now the function (3), for which conditions (11) assume the form

$$\frac{\partial f}{\partial z} = \frac{\partial \varphi_0}{\partial z} + p\varphi_0 \quad \text{for} \quad y = 0, \quad 0 < z < b;$$
$$\frac{\partial^2 f}{\partial y\, \partial z} = 0 \quad \text{for} \quad y = 0, \quad b < z < \infty.$$

Recognizing that the function f vanishes at infinity, we can put

$$f = C_1 + \int \left(\frac{\partial \varphi_0}{\partial z} + p\varphi_0\right) dz = U(z) \quad \text{for} \quad y = 0, \quad 0 < z < b;$$
$$\frac{\partial f}{\partial y} = 0 \quad \text{for} \quad y = 0 \quad b < z < \infty, \tag{12}$$

where C_1 is the constant of integration, which must be determined when the conditions (11) are alternately satisfied. On the basis of (4), the conditions (12) are continued here in even fashion to the lower half plane, and by virtue of the second condition of (12), we have $f(-y, z) = f(y, z)$ and the values of $\partial f/\partial y$ have different signs on opposite sides of the segment (-b, b) of the z axis. Therefore, contracting the contour L + L* in (6) to this segment, we have

$$A(\pm h, p) = \int_{-b}^{b} \frac{\partial f^+}{\partial y} e^{pz} dz \qquad \left(\frac{\partial f^+}{\partial y} = \left(\frac{\partial f}{\partial y}\right)_{y=+0}\right). \tag{13}$$

We put in (5) y = +0, 0 < z < b; then, taking (11) and the properties of the function f into account, we obtain the relation

$$i\frac{p}{h} A(h, p) - \int_b^\infty e^{-pz} \frac{\partial f}{\partial z} dz = \varphi_0(0, b)\, e^{-pb} + 2p \int_0^b e^{-pz} \varphi_0(0, z)\, dz, \tag{14}$$

which determines the constant C_1. We now calculate the densities of the surface currents $i_z^{\pm} = \mp H_x$ on the shadow and illuminated sides, respectively, of a perfectly conducting plane (y = ± 0, z ≥ b), which according to (5) and (11) are determined in the form

$$i_z^+ = -k^2 e^{-pz} \left(\varphi_0(0, b)\, e^{pb} + \int_b^z e^{pz} \frac{\partial f}{\partial z} dz\right),$$
$$i_z^- = 2k^2 \varphi_0(0, z) + i_z^+, \tag{15}$$

and consequently we obtain for the current density the expression

$$i_z = i_z^+ + i_z^- = 2k^2 \left[\varphi_0(0, z) - e^{-pz} \left(\varphi_0(0, b)\, e^{pb} + \int_b^z e^{pz} \frac{\partial f}{\partial z} dz\right)\right]. \tag{16}$$

Let us show that the complex power (see (8))

$$P_k = i\rho_0 k^3 \int_0^b \frac{\partial \varphi^+}{\partial y} \varphi_0^* (0, z)\, dz \tag{17}$$

can be expressed in simple fashion in terms of the function f. In fact, from (5) we have ($y = +0$, $0 < z < b$)

$$\frac{\partial \varphi^+}{\partial y} = \frac{\partial f^+}{\partial y} + pe^{-pz}\left(A(\pm h, p) - \int_{-b}^{z} e^{pz} \frac{\partial f^+}{\partial y}\, dz\right) =$$

$$= \frac{\partial f^+}{\partial y} + pe^{-pz} \int_z^b e^{pz} \frac{\partial f^+}{\partial y}\, dz, \tag{18}$$

and in the case of a function φ_0, determined from (9), we get

$$P_k = i\rho_0 k^3 \frac{ik \sin \beta^*}{ik \sin \beta^* + p} \int_{-b}^{b} \frac{\partial f^+}{\partial y} e^{ikz \sin \beta^*} dz \tag{19}$$

In particular, when $\beta = i\alpha_0$ ($k \operatorname{sh} \alpha_0 = p$), i.e., when the incoming wave is a surface wave, the complex power P_k is expressed in terms of $A(\pm h, p)$ very simply

$$P_k = \frac{i}{2} \rho_0 k^3 A(\pm h, p) = \frac{\rho_0 k^3 h}{2p} D_0, \quad D_0 = \frac{ip}{h} A(h, p), \quad R_0 = 1 - D_0, \tag{20}$$

where D_0 and R_0 are the coefficients of transmission and reflection of the surface waves.

Using (18) we can also readily establish an expression for the total voltage V applied to the edges of the slot:

$$V = \int_0^b E_z (0, z)\, dz = ik\rho_0 \int_0^b \frac{\partial f^+}{\partial y} e^{pz} dz. \tag{21}$$

We note also that for the incident surface waves (10) the equality (18) can be represented in the form

$$\frac{\partial \varphi^+}{\partial y} = -ihe^{-pz} + ihR_0 e^{-pz} + \frac{\partial f^+}{\partial y} - pe^{-pz} \int_{-b}^{z} e^{pz} \frac{\partial f^+}{\partial y}\, dz \quad (y = +0,\ 0 < z < b). \tag{22}$$

The first term in (22) corresponds to the Kirchhoff approximation, and all others take into account the effect of reflection and scattering of the waves.

To determine the function f we introduce a system of elliptical coordinates

$$z = b \operatorname{ch} \xi \cos \eta, \quad y = b \operatorname{sh} \xi \sin \eta. \tag{23}$$

It is easy to see that the coordinate lines $\xi = \text{const}$ correspond to a set of confocal ellipses in the zy plane, while the coordinate lines $\eta = \text{const}$ correspond to a set of confocal hyperbolas orthogonal to the ellipses. The twice-covered segment $(-b, b)$ of the z axis is a degenerate ellipse $\xi = 0$ and $|\eta| < \pi$, and parts of the coordinate axes with $|z| > b$ correspond to the degenerate hyperbolas $\eta = 0$ and $\eta = \pm\pi$ ($\xi > 0$).

In the elliptical system of coordinates, particular solutions of the wave equation are the set of odd and even Mathieu functions $\varphi_n^{(1)} = \mathrm{Ce}_n(\xi)\, \mathrm{ce}_n(\eta)$ ($n = 0, 1, \ldots$) and $\varphi_n^{(2)} = \mathrm{Se}_n(\xi)\, \mathrm{se}_n(\eta)$ ($n = 1, 2, \ldots$), the functions $\mathrm{ce}_n(\eta)$ and $\mathrm{se}_n(\eta)$ being orthogonal systems of periodic functions with a normalization

$$\int_{-\pi}^{\pi} [ce_n(\eta)]^2 d\eta = \pi, \quad \int_{-\pi}^{\pi} [se_n(\eta)]^2 d\eta = \pi.$$

The Mathieu functions $ce_n(\eta)$ and $se_n(\eta)$ are represented by Fourier expansions in the following form [4]:

$$ce_n(\eta) = \sum_{m=0}^{\infty} A_{nm} \cos m\eta, \quad se_n(\eta) = \sum_{m=1}^{\infty} B_{nm} \sin m\eta, \tag{24}$$

where the indices n and m have the same parity, and the coefficients A_{nm} and B_{nm} are entire functions of the parameter $\theta = (kb)^2/4$. The Mathieu-Hankel functions $Ce_n(\xi)$ and $Se_n(\xi)$ are also expressed in terms of series of Bessel and Hankel functions [5, 6].

The function f, satisfying conditions (12), can be determined by series expansion in powers of $\varphi_n^{(1)}$, as a result of which we obtain

$$f = \sum_{n=0}^{\infty} a_n \frac{Ce_n(\xi)}{Ce_n(0)} ce_n(\eta), \quad a_n = \sum_{m=0}^{\infty} A_{nm} d_m^{(1)},$$
$$d_m^{(1)} = \frac{1}{\pi} \int_{-\pi}^{\pi} U(b \cos\eta) \cos m\eta \, d\eta. \tag{25}$$

For the function φ_0, determined from (9), the coefficients $d_m^{(1)}$ are easy to calculate. In this case

$$U(z) = C_1 + \left(1 + \frac{p}{ik \sin\beta}\right)(e^{ikz \sin\beta} + e^{-ikz \sin\beta}). \tag{26}$$

Therefore, using the Bessel integral

$$(\pm i)^m J_m(x) = \frac{1}{\pi} \int_0^{\pi} e^{\pm ix \cos\eta} \cos m\eta \, d\eta, \tag{27}$$

we obtain ($\delta_0 = 1$, $\delta_m = 0$ when $m \neq 0$)

$$d_m^{(1)} = 2\delta_m C_1 + 2i^m (1 + (-1)^m)\left(1 + \frac{p}{ik \sin\beta}\right) J_m(kb \sin\beta). \tag{28}$$

In precisely the same manner we can find the complex amplitude $A(\pm h, p)$ of the surface waves. From (25) we determine

$$\frac{\partial f^+}{\partial y} = \frac{1}{b \sin\eta} \sum_{n=0}^{\infty} a_n \frac{Ce_n'(0)}{Ce_n(0)} ce_n(\eta) \quad \left(Ce_n'(0) = \left(\frac{dCe_n(\xi)}{d\xi}\right)_{\xi=0}\right). \tag{29}$$

Substituting this expression in (13), we get

$$A(\pm h, p) = \pi \sum_{n=0}^{\infty} a_n \frac{Ce_n'(0)}{Ce_n(0)} D_n^{(1)}(\nu), \quad D_n^{(1)}(\nu) = \sum_{m=0}^{\infty} A_{nm} I_m(\nu) \qquad (\nu = pb). \tag{30}$$

Here $I_m(\nu)$ is the Bessel function of imaginary argument

$$I_m(\nu) = \frac{1}{\pi} \int_0^{\pi} e^{\nu \cos\eta} \cos m\eta \, d\eta. \tag{31}$$

We express in similar form the complex power P_k and the voltage V. Actually, we can establish from the formulas (19), (21), and (29) that ($\alpha_0^{(1)} = 0$, $\alpha_m^{(1)} = 1/m$)

$$P_k = \pi i \rho_0 k^3 \frac{ik \sin \beta^*}{ik \sin \beta^* + \gamma} \sum_{n=0}^{\infty} a_n \frac{Ce_n'(0)}{Ce_n(0)} D_n^{(1)} (ikb \sin \beta^*), \tag{32}$$

$$V = ik\rho \sum_{n=0}^{\infty} a_n \frac{Ce_n'(0)}{Ce_n(0)} B_n, \qquad B_n = \sum_{m=0}^{\infty} A_{nm} \left(\nu^m S_m(\nu) + \alpha_m^{(1)} \sin \frac{m\pi}{2} \right), \tag{33}$$

where $S_m(\nu)$ stands for the following functions

$$S_m(\nu) = \frac{1}{\nu^m} \left(\int_0^{\pi/2} e^{\nu \cos \eta} \cos m\eta \, d\eta - \alpha_m^{(1)} \sin \frac{m\pi}{2} \right), \tag{34}$$

for which we obtain, integrating by parts, the following recurrence relation

$$S_{m+1}(\nu) = \frac{1}{\nu^2} S_{m-1}(\nu) - \frac{2m}{\nu^2} S_m(\nu) - \frac{1}{\nu^{m+1}} (\alpha_{m-1}^{(1)} + \alpha_{m+1}^{(1)}) \cos \frac{m\pi}{2}$$
$$(m = 1, 2, \ldots). \tag{35}$$

This relation determines all the functions $S_m(\nu)$ in terms of $S_0(\nu)$ and $S_1(\nu)$. The latter has been investigated in [7] and are expressed in terms of the Bessel functions and the zero-and first-order Struve functions $L_0(\nu)$ and $L_1(\nu)$ of imaginary argument [8], in the following manner:

$$S_0(\nu) = \frac{\pi}{2} \left(I_0(\nu) + L_0(\nu) \right), \qquad S_1(\nu) = \frac{\pi}{2\nu} \left(I_1(\nu) + L_1(\nu) \right), \tag{36}$$

with the following identities derived for the Struve functions [7]

$$L_0(\nu) = -\frac{2}{\pi} \left[I_0^{-1}(\nu) K_0(\nu) - K_0^{-1}(\nu) I_0(\nu) \right], \qquad I_0^{-1}(\nu) = \int_0^{\nu} I_0(\nu) \, d\nu,$$

$$L_1(\nu) = \frac{2}{\pi} \left[I_0^{-1}(\nu) K_1(\nu) + K_0^{-1}(\nu) I_1(\nu) - 1 \right], \qquad K_0^{-1}(\nu) = \int_0^{\nu} K_0(\nu) \, d\nu,$$

where $K_n(\nu)$ are modified Hankel functions of imaginary argument (Macdonald functions).

To determine the constant C_1 contained in the solution (25) it is necessary to satisfy condition (14). The integral term in this condition is best calculated with the aid of equation

$$f = \frac{i}{2} \int_{-b}^{b} \frac{\partial f^+}{\partial y} H_0^{(2)} [k (y^2 + (z - s)^{1/2}] \, ds, \tag{37}$$

in accordance with which we get

$$\int_b^{\infty} e^{-pz} \frac{\partial f}{\partial z} dz = \frac{i}{2} \int_{-b}^{b} \frac{\partial f^+}{\partial y} (p Q(s) - H_0^{(2)}(k(b - s))) \, ds. \tag{38}$$

$$Q(s) = e^{-ps} \int_{b-s}^{\infty} e^{-pt} H_0^{(2)}(kt) \, dt. \tag{39}$$

The procedure for further calculating the individual coefficients in (14) is approximately the same as in problems involving the vibrations of thin wings in subsonic streams [9], and we therefore omit the details of the calculations.

As is well known, the solution (25), which we obtained in the form of an expansion in Mathieu functions, can be used in calculations for values $kb \leq 1$ and $kb \sim 1$, and is inconvenient in practice for values $kb \gg 1$. On the other hand, when $kb \ll 1$ the limiting solution (25) assumes a simpler form. When $kb \ll 1$ we have $\varphi_n^{(1)} \simeq \cos n\eta \exp(-n\xi)$ and $\varphi_n^{(2)} \simeq \sin n\eta \exp(-n\xi)$, with $\varphi_0^{(1)} \simeq \xi$, and all the calculations become much simpler. Considering, in particular, the diffraction of surface waves on a slot, i.e., when $\beta = i\alpha_0 (p = k \operatorname{sh} \alpha_0)$, we get in this case $\varphi_0(0, z) = \exp(-pz)$ and $U(z) = \text{const}$. Therefore condition (14) becomes simpler

$$i\frac{p}{h}A - \int_b^\infty e^{-pz}\frac{\partial f}{\partial z}dz = 1, \tag{40}$$

where when $kb \ll 1$ the function f is determined in effective form

$$f = a_0\xi + \text{const}, \tag{41}$$

in which the inessential constant can be discarded. At the point of the z axis we have

$$\frac{\partial f}{\partial y} = \frac{a_0}{b \sin \eta}, \quad \xi = 0, \quad 0 < \eta < \pi; \quad \frac{\partial f}{\partial z} = \frac{a_0}{b \operatorname{sh} \xi}, \quad \eta = 0, \quad \xi > 0. \tag{42}$$

On the basis of (13) we obtain

$$A(\pm h, p) = \pi a_0 I_0(\nu) \quad (\nu = pb) \tag{43}$$

Taking also account of the equality

$$\int_0^\infty e^{-\nu \operatorname{ch} \xi} d\xi = K_0(\nu) \tag{44}$$

and using (40), (42), and (43), we obtain

$$a_0 = \frac{1}{\pi i p h^{-1} I_0(\nu) - K_0(\nu)}, \qquad A(\pm h, p) = \frac{-\pi i I_0(\nu)}{\pi p h^{-1} I_0(\nu) + i K_0(\nu)} \tag{45}$$

Let us separate the active and reactive powers in Eq. (20), assuming $P_k = 2P_a + i2P_r$; we can then readily establish that

$$P_a = P_+ = \frac{1}{4}\rho_0 k^3 p h^{-1} \frac{(\pi I_0(\nu))^2}{\Delta_1^2}, \qquad P_r = -\frac{1}{4}\rho_0 k^3 \frac{\pi I_0(\nu) K_0(\nu)}{\Delta_1^2} \tag{46}$$

$$(\Delta_1^2 = (\pi p h^{-1} I_0(\nu))^2 + K_0^2(\nu)),$$

where P_+ is determined from (7) and (45), so that the total active power is consumed entirely in the production of surface waves. This result is not unexpected, for when $kb \ll 1$ only the surface waves are radiated, and no cylindrical waves are radiated. Furthermore, we get from (21), (34) and (42)

$$V = -\rho_0 k \frac{S_0(\nu)}{\pi p h^{-1} I_0(\nu) + i K_0(\nu)}. \tag{47}$$

Consequently, the complex radiation admittance is determined by the formula

$$g_k = \frac{P_k}{|V|^2} = \frac{k}{2\rho_0 S_0^2(\nu)} [p h^{-1} (\pi I_0(\nu))^2 - \pi i I_0(\nu) K_0(\nu)]. \tag{48}$$

Let us stop to discuss in greater detail the numerical values of transmission coefficient of the surface waves:

$$D_0 = iph^{-1}A\,(h, p) = |D_0|\,e^{-i\theta_0}, \qquad |D_0| = \frac{\pi ph^{-1}I_0(\nu)}{\Delta_1},$$
$$\operatorname{tg}\theta_0 = \frac{K_0(\nu)}{\pi ph^{-1}I_0(\nu)}, \qquad ph^{-1} = (1 - \beta_0^2)^{1/2}, \qquad \beta_0 = \frac{k}{h}. \tag{49}$$

Here β_0 is the slowing-down coefficient. It must be kept in mind, however, that we have assumed that pb/hb $\simeq 1$, and therefore this approximation in effect disregards the slowing down and can be used for estimating purposes only when β_0 is small. Figs. 2 and 3 show the plots of $|D_0|$ and θ_0 vs. ν when $\beta_0 = 0$ and $\beta_0 = 0.6$. It follows from calculations ($\beta_0 = 0$) that when $\nu = 0.02$ we get $|D_0| = 0.61$ and $\theta_0 = 52°$, and when $\nu = 0.4$ we get $|D_0| = 0.95$ and $\theta_0 = 19°$. To discuss this data we introduce the effective height $\Delta = 1/p$

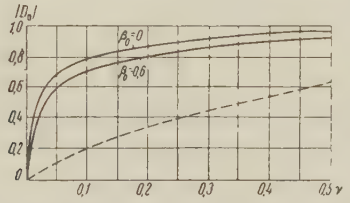


Fig. 2

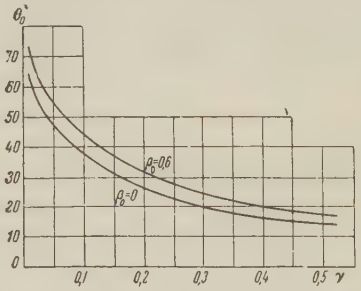


Fig. 3

of the zone of action of the surface waves; it turns out that when $\Delta \leq 2.5$ b all the surface waves are transmitted and none are reflected. This means that the surface waves are capable of penetrating through the narrow slot (kb $\ll 1$) and are only slightly distorted when $\Delta \leq 2.5$ b. One might therefore think that in relation (22) the predominant role is assumed by the first term (the Kirchhoff approximation). On the other hand, the transmission coefficient can be calculated with the aid of the exact formula [3]

$$D_0 = 2i\frac{p}{h}\int_0^b \frac{\partial\varphi^+}{\partial y}e^{-pz}\,dz. \tag{50}$$

If we not use the Kirchhoff approximation, putting $\partial\varphi^+/\partial y \simeq -ih\exp(-pz)$, then we obtain from (50)

$$D_0 \simeq 1 - e^{-2\nu}, \tag{51}$$

which is represented by the dashed curve in Fig. 2; this leads to values that are by far too low, which is quite understandable, since this is the least favorable case for the approximation considered [3].

Let us also consider the expression for the relative value of surface current $i_z^0 = i_z/k^2$ determined in accordance with (16) and (24) in the form

$$i_z^0 = -2a_0e^{-\nu z_0}\int_1^{z_0}\frac{e^{\nu t}\,dt}{(t^2-1)^{1/2}} \tag{52}$$
$$(z_0 = z/b).$$

Fig. 4 shows a plot of $|i_z^0|$ vs $x = p(z-b)$ when $\beta_0 = 0$. The same plot shows the limiting curve $i_z^0 = 2\exp(-pz)$ (pb = 0) corresponding to total reflection of the incident surface waves.

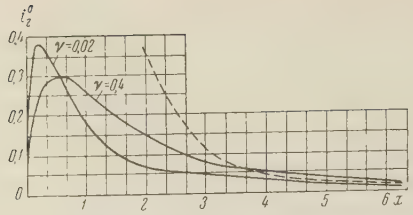


Fig. 4

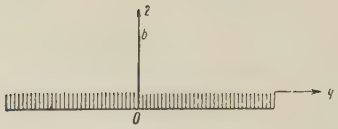


Fig. 5

3. DIFFRACTION OF WAVES ON A RIBBON

Assume that an arbitrary specified field, characterized by a function $\varphi_0(y, z)$, strikes a perfectly conducting ribbon, placed perpendicular to the impedance plane (Fig. 5). In this case it is convenient to specify the total field over the entire upper half-plane with the aid of the function $\varphi + \varphi_0$, where the scalar function φ determines the field of the scattered waves. In such a formulation the boundary condition on a perfectly conducting ribbon has the form

$$\frac{\partial \varphi}{\partial y} = -\frac{\partial \varphi_0}{\partial y} \quad \text{for } y = 0, \quad 0 < z < b, \tag{53}$$

and this condition for the function (3) yields

$$\frac{\partial f}{\partial y} = C_2 - \int \left(\frac{\partial}{\partial z} + p\right) \frac{\partial \varphi_0}{\partial y} dz = U(z) \quad \text{for } y = 0,\ 0 < z < b, \tag{54}$$

where the constant C_2 must be determined when the condition (53) is ultimately satisfied. In addition, on the basis of (4) the values of $\partial f/\partial y$ continue into the lower half-plane in even fashion. For the function φ_0, determined from (9) we obtain

$$U(z) = C_2 + \operatorname{ctg} \beta\ (ik \sin \beta + + p)\ (e^{ikz \sin \beta} + e^{-ikz \sin \beta}). \tag{55}$$

From symmetry considerations it is obvious that $f(-y, z) = -f(y, z)$. Therefore the values of the function f on both sides of the segment (-b, b) of the z axis are opposite in sign, and the function f vanishes on the remaining parts of the z axis. Taking these properties into account and contracting in (6) the contour L + L* to the segment (-b, b), we obtain

$$A(\pm h, p) = \mp ih \int_{-b}^{b} f^+ e^{pz} dz = \pm i \frac{h}{p} \int_{-b}^{b} \frac{\partial f^+}{\partial z} e^{pz} dz \quad (f^+ = f(+0, z)). \tag{56}$$

Differentiating (5) with respect to y and setting y = 0, we get

$$\frac{\partial \varphi}{\partial y} = \frac{\partial f}{\partial y} - pe^{-pz} \int_{-\infty}^{z} e^{pz} \frac{\partial f}{\partial y} dz + pA(h, p) e^{pz}. \tag{57}$$

We now satisfy condition (53) and arrive at

$$p \int_{b}^{\infty} e^{-pz} \frac{\partial f}{\partial y} dz - pA(h, p) = e^{-pb} \left(\frac{\partial \varphi_0}{\partial y} + U\right)_{z=b} + 2p \int_{0}^{b} e^{-pz} \frac{\partial \varphi_0}{\partial y} dz, \tag{58}$$

which determines the constant C_2. For incident surface waves (10) we have $U = C_2$ and (57) simplifies somewhat

$$p \int_{b}^{\infty} e^{-pz} \frac{\partial f}{\partial y} dz - pA(h, p) = C_2 e^{-pb} - ih, \tag{59}$$

while the transmission and reflection coefficients of the surface waves are connected with A(± h, p) by the relation

$$D_0 = 1 \pm i\frac{p}{h}A(h, p), \quad R_0 = i\frac{p}{h}A(-h, p) = 1 - D_0. \tag{60}$$

In this case formula (20) remains in force for the complex power P_k, calculated similarly, and the incident waves must be taken into account in calculating the mean power P_+ carried by the surface waves as $y \to \infty$; we then obtain

$$P_+ = \frac{\rho_0 k^3 h}{4p}|D_0|^2. \tag{61}$$

On the basis of (5) and (56) we readily determine the surface density and the total current flowing in the ribbon

$$i_z = -2k^2 e^{-pz}\int_b^z e^{pz}\frac{df^+}{\partial z}dz,$$

$$I_z = \int_0^b i_z dz = -\frac{2k^2}{p}\int_0^b \frac{\partial f^+}{\partial z}(1 - e^{pz})\,dz. \tag{62}$$

We note also that for the function (10) we can obtain a simple expression for the voltage V between the upper edge of the ribbon and an infinitely remote point:

$$V = ik\rho_0\int_b^\infty \frac{\partial(\varphi + \varphi_0)}{\partial y}dz = \rho_0\frac{ik}{p}C_2.$$

The solution of this problem is constructed in the same manner as before, by expanding f in powers of odd Mathieu functions

$$f = \sum_{n=1}^{\infty} b_n \frac{Se_n(\xi)}{Se_n'(0)} se_n(\eta) \quad \left(Se_n'(0) = \left(\frac{dSe_n(\xi)}{d\xi}\right)_{\xi=0}\right). \tag{63}$$

This relation satisfies the condition f(0, z) = 0 when $|z| > b$. Satisfying in addition the condition (54), we obtain expressions for the coefficients b_n

$$b_n = \sum_{m=1}^{\infty} B_{nm} d_m^{(2)}, \quad d_m^{(2)} = \frac{2b}{\pi}\int_0^\pi \sin\eta \sin m\eta U(b\cos\eta)\,d\eta, \tag{64}$$

and for the function (55) we determine $d_m^{(2)}$ in terms of the Bessel functions ($\delta_1 = 1$, $\delta_m = 0$ when $m \neq 1$):

$$d_m^{(2)} = b\left[\delta_m C_2 + 2m\operatorname{ctg}\beta(1 - (-1)^m)\left(1 + \frac{p}{ik\sin\beta}\right)i^m J_m(kb\sin\beta)\right]. \tag{65}$$

It is easy to establish general expressions for A(± h, p) and I_z. For this purpose we take into account (56), (62), (63), (31), and (34) and obtain ultimately

$$A(\pm h, p) = \mp \pi i\frac{h}{p}\sum_{n=1}^{\infty}\frac{Se_n(0)}{Se_n'(0)}b_n D_n^{(2)}, \quad D_n^{(2)} = \sum_{m=1}^{\infty} mB_{nm}I_m \tag{66}$$

$$I_z = -\frac{2k^2}{p}\sum_{n=1}^{\infty}\frac{Se_n(0)}{Se_n'(0)}b_n Y_n, \quad Y_n = \sum_{m=1}^{\infty} B_{nm}m\nu^m S_m(\nu) \quad (\nu = pb). \tag{67}$$

For the final solution we must determine C_2 with the aid of (58). The first integral term on the left side is best transformed first with the aid of the equation

$$f = \frac{i}{2}\frac{\partial}{\partial y}\int_{-b}^{b} f^{+} H_0^{(2)}\left[(k(z-s)^2 + y^2)^{1/2}\right] ds; \tag{68}$$

so that we get

$$\int_b^\infty e^{-pz}\frac{\partial f}{\partial y}\,dz = \frac{i}{2}\int_{-b}^{b} f^{+}\left[e^{-pb}\left(pH_0^{(2)}(k(b-s)) - kH_1^{(2)}(k(b-s)) - h^2 Q(s)\right)\right] ds, \tag{69}$$

where the function Q(s) is determined from (39).

Let us investigate the limiting solution for $kb \ll 1$ in the simplest case, that of incident surface waves (10). The function f is determined here simply as

$$f = -bC_2 e^{-\xi}\sin\eta. \tag{70}$$

From (56) and (59) we obtain directly

$$A(\pm h, p) = \pm\pi i\frac{h}{p}C_2 I_1(\nu), \quad C_2 = \frac{1}{b}\frac{i}{ph^{-1}K_1(\nu) + \pi i I_1(\nu)}; \tag{71}$$

and therefore, in accordance with (60), we have for the reflection and transmission coefficients

$$R_0 = \frac{\pi I_1(\nu)}{\pi I_1(\nu) - iph^{-1}K_1(\nu)},$$
$$D_0 = \frac{ph^{-1}K_1(\nu)}{ph^{-1}K_1(\nu) + \pi i I_1(\nu)} = |D_0| e^{-i\theta_0}\left(\operatorname{tg}\theta_0 = \frac{\pi I_1(\nu)}{ph^{-1}K_1(\nu)}\right), \tag{72}$$

with $P_a = P_+$, since $\mathrm{Re}D_0 = |D_0|^2$ and is brought about by the same factors as in the case considered earlier.

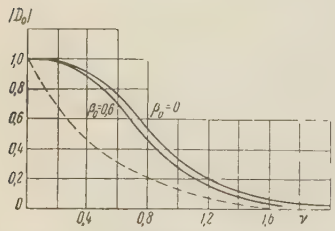


Fig. 6

Figs. 6 and 7 show the dependences of $|D_0|$ and θ_0 on ν. We see that when $\nu \simeq 1.5$ the surface-wave transmission coefficient vanishes, and consequently total reflection of the surface waves, without transmission, occurs only when the effective height $\Delta = 1/p$ is not greater than 2/3 of the width of the ribbon.

Let us compare this data with the Kirchoff approximation, which like (50) is determined from the exact formula

$$D_0 = 2i\frac{p}{h}\int_b^\infty \frac{\partial\varphi^{+}}{\partial y}e^{-pz}\,dz \simeq 2i\frac{p}{h}\int_b^\infty \frac{\partial\varphi_0}{\partial y}e^{-pz}dz = e^{-2\nu}. \tag{73}$$

The relationship (73) is represented in Fig. 6 by the dashed line.

We now substitute (70) into (62), and obtain for the total current

$$I_z = I_0 bC_2 S_1(\nu) = |I_z| e^{i\left(\frac{\pi}{2} - \theta_0\right)} \quad (I_0 = 2k^2 b). \tag{74}$$

Fig. 8 shows a plot of $|I_z|/I_0$ ($\beta_0 = 0$) and the limiting curve $I_z/I_0 = \nu^{-1}(1 - \exp(-\nu))$ corresponding to total reflection of the incident waves (dashed line). It follows from the figure that $\max|I_z| = 0.695\, I_0$ is reached when $\nu = 0.78$.

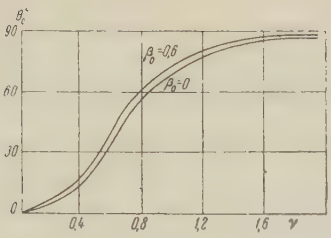


Fig. 7

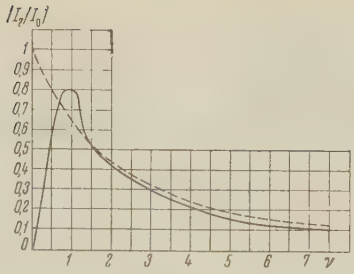


Fig. 8

In conclusion we note that the exact solution developed here for a ribbon generalizes, from the mathematical point of view, one of the solutions in reference [10], while the limiting problems concerning diffraction of surface waves on a slit and on a ribbon at kb $\ll$ 1 correspond to the hydrodynamic problems [11, 7, 12] investigated by a different method.

REFERENCES

1. M.D. Khaskind. Propagation of acoustic and electromagnetic waves in a half-space. Akust. zh., 1959, 5, 4, 464.
2. M.D. Khaskind. Excitation of surface electromagnetic waves on plane dielectric coatings. Radiotekhnika i elektronika. 1960, 5, 2, 188.
3. M.D. Khaskind. Excitation of waves on an impedance plane. Ibid, 1961, 6, 8, 1259.
4. M.J.O. Strutt. Lame, Mathieu, and related functions in Physics and engineering (in German). Edwards Bros. Ann Arbor, Mich, 1935.
5. V.D. Kupradze. Fundamental problems in the mathematical theory of diffraction (steady state processes). ONTI, 1935.
6. N.W. McLachlan. Theory and application of Mathieu functions. Clarendon Press, Oxford, 1947.
7. M.D. Khaskind. Radiation and diffraction of surface waves by a vertical floating plate. Prikladnaya matematika i mekhanika, 1959, 23, 3, 546.
8. G.N. Watson. A treatise on the theory of Bessel functions. Cambridge University Press, 1944.
9. G. Blanch, H.E. Fettis. Subsonic oscillatory aerodynamic coefficients computed by the method of Reissner and Haskind. J. Aeronaut. Sci., 1953, 20, 12, 851.
10. M.D. Khaskind. Diffraction of waves around a moving cylindrical vessel. Prikladnaya matematika i mekhanika. 1953, 17, 4, 431.
11. M.D. Khaskind. Pressure of waves on a partition. Inzhenernyy sbornik Instituta mekhanika, AN SSSR, 1948, 4, 2, 147.
12. I.A. Mariyanskiy. Diffraction of waves around an underwater vertical plate. Prikladnaya matematika i mekhanika, 1954, 17, 2, 233.

Received by editor 21 November 1960

DIELECTRIC WAVEGUIDE WITH ELLIPTICAL CROSS SECTION

L.A. Lyubimov, G.I. Veselov, and N.A. Bei

The principles of the theory of a dielectric waveguide of elliptical cross section are developed. The analysis is limited to a derivation of the dispersion equations and of the critical conditions. Results of calculations of the dispersion curves for the main modes ('even' and 'odd') and critical dimensions with respect to the nearest higher modes are derived in a form that is comparable with the case of a round cross section.

INTRODUCTION

Dielectric waveguide behavior is exhibited by many waveguide devices containing dielectric plates. Examples of such devices are phase shifters, polarizers, and units with ferrites. The understanding and a quantitative estimate of the phenomena occurring in these devices are hindered by the lack of a theory of the dielectric waveguide of elliptical cross section, for in most cases dielectrics of drawn (rectangular) section are used. Many complex phenomena have been investigated in the papers available to us on the theory of the dielectric waveguide [1, 2, 3, and others], but the dielectrics considered are always round. The case of elliptic cross section necessitates many cumbersome calculations using Mathieu functions, and also the use of sensible simplifications. The calculations and simplifications made enabled us to obtain many important relationships, including dispersion equations in relatively simple form, amenable to further treatment.

1. ELECTROMAGNETIC FIELD COMPONENTS

An analysis of the wave processes occurring inside and near a dielectric rod of elliptic cross section (Fig. 1) is made by using a system of 'elliptic cylinder' coordinates (ξ, η, z). The coordinates ξ and η are connected with the rectangular coordinates x and y by relations

$$x = c \operatorname{ch} \xi \cos \eta,$$
$$y = c \operatorname{sh} \xi \sin \eta.$$

The parameter ξ determines the degree of ellipticity of the cylinder, and the quantities $c \operatorname{ch} \xi$ and $s \operatorname{sh} \xi$ give the characteristic dimensions of the cross section (semi-axes). The law of variation of the field components along the z axis is assumed for the forward wave to have the form $e^{-\gamma z}$, where γ is the propagation constant. The wave factor $\exp[j(\omega t - \gamma z)]$ is omitted throughout.

We take as the initial quantities the longitudinal components of the field E_z and H_z. The wave equations

$$\Delta E_z + \omega^2 \varepsilon \mu E_z = 0,$$
$$\Delta H_z + \omega^2 \varepsilon \mu H_z = 0$$

assume in the case of the elliptic-cylinder coordinates the following form

$$\frac{\partial^2 E_z}{\partial \xi^2} + \frac{\partial^2 E_z}{\partial \eta^2} = -E_z \frac{\vartheta^2}{2} [\operatorname{ch} 2\xi - \cos 2\eta], \tag{1}$$

where $\theta^2 = K_c^2 c^2$; $K_c^2 = \gamma^2 + \omega^2 \varepsilon \mu$; c — parameter of coordinate transformation.

The solution (1), sought in the form $E_z = F(\xi)\,H(\eta)$, leads to the equations

$$\frac{\partial^2 H}{\partial \eta^2} + \left(\lambda - \frac{\vartheta^2}{2}\cos 2\eta\right) H = 0,$$

$$\frac{\partial^2 F}{\partial \xi^2} + \left(-\lambda + \frac{\vartheta^2}{2}\operatorname{ch} 2\xi\right) F = 0.$$

The constant λ should have a value such as to yield, for a specified ϑ^2, periodic solutions of the Mathieu equation for the function $H(\eta)$. The numbers of the parameter λ for each ϑ^2 will yield the numbers of the wave harmonics (with respect to the parameter η.

Further analysis is made difficult partly because of having to use the untabulated radial Matheiu functions, but principally because of the difficulty in formulating the boundary conditions. The point is that the parameters K_{c1}^2 and K_{c2}^2 (for the internal and external regions of the dielectric waveguide) differ in magnitude and in sign ($K_{c1}^2 > 0$, $K_{c2}^2 < 0$). The parameters of the wave equations ϑ_1^2 and ϑ_2^2 are accordingly also different. Because of this the angular functions $C_n(\vartheta, \eta)$ and $S_n(\vartheta, \eta)$ are not the same for two regions on the common boundary ξ_0 = const (Fig. 1). In other words, in order to equate identically the tangential components on the common boundary ξ_0 = const, it becomes necessary to use series of angle functions, which makes the problem very cumbersome as a whole.

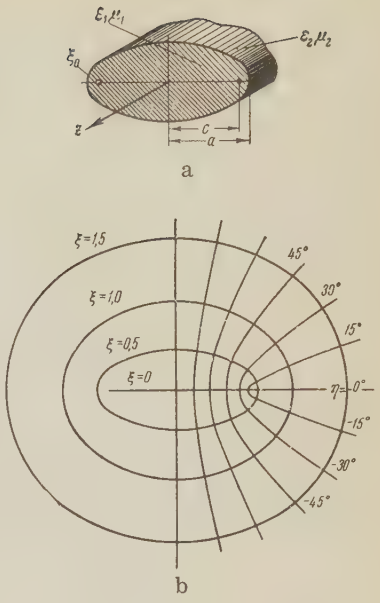


Fig. 1. Shape of dielectric waveguide (a) and coordinate system employed (b)

For the outer region we obtain the following expressions for the components:

$$E_{z,e} = \sum_{\nu=1}^{\infty} A_\nu Rc_\nu(\vartheta_2, \xi)\, C_\nu(\vartheta_2, \eta) + \sum_{\nu=1}^{\infty} C_\nu Rs_\nu(\vartheta_2, \xi)\, S_\nu(\vartheta_2, \eta), \tag{2}$$

$$H_{z,e} = \sum_{\nu=1}^{\infty} B_\nu Rc_\nu(\vartheta_2, \xi)\, C_\nu(\vartheta_2, \eta) + \sum_{\nu=1}^{\infty} D_\nu Rs_\nu(\vartheta_2, \xi)\, S_\nu(\vartheta_2, \eta), \tag{3}$$

$$E_{\xi,e} = -\frac{1}{K_{c_2}^2 c\sqrt{\operatorname{ch}^2\xi - \cos^2\eta}}\Big[\gamma\sum_{\nu=1}^{\infty} A_\nu R'c_\nu(\vartheta_2, \xi)\, C_\nu(\vartheta_2, \eta) +$$

$$+ \gamma\sum_{\nu=1}^{\infty} C_\nu R's_\nu(\vartheta_2, \xi)\, S_\nu(\vartheta_2, \eta) +$$

$$- j\omega\mu_2\sum_{\nu=1}^{\infty} B_\nu Rc_\nu(\vartheta_2, \xi)\, C'_\nu(\vartheta_2, \eta) + j\omega\mu_2\sum_{\nu=1}^{\infty} D_\nu Rs_\nu(\vartheta_2, \xi)\, S'_\nu(\vartheta_2, \eta)\Big], \tag{4}$$

$$E_{\eta,e} = -\frac{1}{K_{c_2}^2 c\sqrt{\operatorname{ch}\xi - \cos^2\eta}}\Big[\gamma\sum_{\nu=1}^{\infty} A_\nu Rc_\nu(\vartheta_2, \xi)\, C'_\nu(\vartheta_2, \eta) +$$

$$+ \gamma\sum_{\nu=1}^{\infty} C_\nu Rs_\nu(\vartheta_2, \xi) S'_\nu(\vartheta_2, \eta) -$$

$$- j\omega\mu_2\sum_{\nu=1}^{\infty} B_\nu R'c_\nu(\vartheta_2, \xi)\, C_\nu(\vartheta_2, \eta) - j\omega\mu_2\sum_{\nu=1}^{\infty} D_\nu R's_\nu(\vartheta_2, \xi)\, S_\nu(\vartheta_2, \eta)\Big], \tag{5}$$

$$H_{\xi,e} = -\frac{1}{k_{c_2}^2 c \sqrt{\operatorname{ch}^2 \xi - \cos^2 \eta}} \Big[- j\omega\varepsilon_2 \sum_{\nu=1}^{\infty} A_\nu Rc_\nu (\vartheta_2, \xi)\, C'_\nu (\vartheta_2, \eta) -$$

$$- j\omega\varepsilon_2 \sum_{\nu=1}^{\infty} C_\nu Rs_\nu (\vartheta_2, \xi)\, S'_\nu (\vartheta_2, \eta) +$$

$$+ \gamma \sum_{\nu=1}^{\infty} B_\nu R'c_\nu (\vartheta_2, \xi)\, C_\nu (\vartheta_2, \eta) + \gamma \sum_{\nu=1}^{\infty} D_\nu R's_\nu (\vartheta_2, \xi)\, S_\nu (\vartheta_2, \eta) \Big], \tag{6}$$

$$H_{\eta,e} = -\frac{1}{k_{c_2}^2 c \sqrt{\operatorname{ch}^2 \xi - \cos^2 \eta}} \Big[j\omega\varepsilon_2 \sum_{\nu=1}^{\infty} A_\nu R'c_\nu (\vartheta_2, \xi)\, C_\nu (\vartheta_2, \eta) +$$

$$- j\omega\varepsilon_2 \sum_{\nu=1}^{\infty} C_\nu R's_\nu (\vartheta_2, \xi)\, S_\nu (\vartheta_2, \eta) + \gamma \sum_{\nu=1}^{\infty} B_\nu Rc_\nu (\vartheta_2, \xi)\, C'_\nu (\vartheta_2, \eta) +$$

$$+ \gamma \sum_{\nu=1}^{\infty} D_\nu Rs_\nu (\vartheta_2, \xi)\, S'_\nu (\vartheta_2, \eta) \Big], \tag{7}$$

where Rc_ν, Rs_ν, C_ν, S_ν — radial and angular Mathieu functions (even and odd) of order ν.

In the internal region the components are also expressed by the series

$$E_{z,i} = \sum_{k=1}^{\infty} A'_k Rc_k (\vartheta_1, \xi)\, C_k (\vartheta_1, \eta) + \sum_{k=1}^{\infty} C'_k Rs_k (\vartheta_1, \xi)\, S_k (\vartheta_1, \eta), \tag{8}$$

$$H_{z,i} = \sum_{k=1}^{\infty} B'_k Rc_k (\vartheta_1, \xi)\, C_k (\vartheta_1, \eta) + \sum_{k=1}^{\infty} D'_k Rs_k (\vartheta_1, \xi)\, S_k (\vartheta_1, \eta). \tag{9}$$

The transverse components are expressed by formulas similar to (4) — (7).

2. BOUNDARY CONDITIONS

The boundary conditions can be written in the following form (with $\xi = \xi_0$)

$$E_{z,i} = E_{z,e},\ H_{z,i} = H_{z,e},\ E_{\eta,e} = E_{\eta,i},\ H_{\eta,e} = H_{\eta,i}. \tag{10}$$

In examining (10) together with (2) — (7), we notice the following. The first two relations in (10) actually relate with one another the coefficient A'_k and A_ν, C'_k and C_ν, B'_k and B_ν, D'_k and D_ν. The third and fourth equations in (10) contain the even and odd functions of the parameter η, which when separated yield separate equations connecting A_ν, D_ν, A'_k with D'_k and C_ν, B_ν, C'_k with B'_k.

Thus, the waves described by the harmonics with coefficients A, A' and D, D' are independent of the waves with harmonic coefficients C, C' and D, D'. The harmonic series can also be grouped by the symmetry conditions with respect to the parity of the numbers ν and k.

The wave types considered below have both electric and longitudinal components, one of which is an even and the other an odd function of the parameter η. In this connection, the distinction between even and odd types, assumed for waves in a metallic waveguide of elliptic cross section [4], will in our case be arbitrary. In order to retain their inherent advantages, it is nevertheless advisable to classify as even the waves with components

$$E_{z,e} = \sum_{\nu=1}^{\infty} C_\nu Rs_\nu (\vartheta_2, \xi)\, S_\nu (\vartheta_2, \eta),$$

$$H_{z,e} = \sum_{\nu=1}^{\infty} B_\nu Rc_\nu (\vartheta_2, \xi)\, C_\nu (\vartheta_2, \eta)$$

and as odd the waves with components

$$E_{z,e} = \sum_{\nu=1}^{\infty} A_\nu Rc_\nu(\vartheta_2, \xi) C_\nu(\vartheta_2, \eta),$$

$$H_{z,e} = \sum_{\nu=1}^{\infty} D_\nu Rs_\nu(\vartheta_2, \xi) S_\nu(\vartheta_2, \eta).$$

Using odd waves as an example, we write out the boundary conditions (10) in expanded form. The first conditions assume the form

$$\sum_\nu A_\nu Rc_\nu(\vartheta_2, \xi_0) C_\nu(\vartheta_2, \eta) = \sum_\nu A_\nu \sum_k A_{\nu,k} Rc_k(\vartheta_1, \xi_0) C_k(\vartheta_1, \eta), \tag{11}$$

$$\sum_\mu D_\mu Rs_\mu(\vartheta_2, \xi_0) S_\mu(\vartheta_2, \eta) = \sum_\mu D_\mu \sum_k D_{\mu,k} Rs_k(\vartheta_1, \xi_0) S_k(\vartheta_1, \eta), \tag{12}$$

where $A_{\nu,k}$ and $D_{\mu,k}$ are the coefficients of expansion of the functions $C_\nu(\theta, \eta)$ and $S_\nu(\theta_2, \eta)$ in functions of the parameter θ_1, multiplied respectively by $Rc_\nu(\theta_2, \xi_0)/Rc_k(\theta_1, \xi_0)$ and $Rs_\mu(\theta_2, \xi_0) / Rs_k(\theta_1, \xi_0)$.

The third and fourth conditions can be recast in the form

$$\gamma\left(\frac{1}{K_{c_1}^2} - \frac{1}{K_{c_2}^2}\right) \sum_\nu A_\nu Rc_\nu(\vartheta_1, \xi_0) C'_\nu(\vartheta_2, \eta) = \frac{j\omega\mu_1}{K_{c_1}^2} \sum_\mu D_\mu \sum_k D_{\mu,k} R's_k(\vartheta_1, \xi_0) \times$$
$$\times S_k(\vartheta_1, \eta) - \frac{j\omega\mu_2}{K_{c_2}^2} \sum_\mu D_\mu \frac{R's_\mu(\vartheta_2, \xi_0)}{Rs_\mu(\vartheta_2, \xi_0)} \sum_k D_{\mu,k} Rs_k(\vartheta_1, \xi_0) S_k(\vartheta_1, \eta), \tag{13}$$

$$\gamma\left(\frac{1}{K_{c_1}^2} - \frac{1}{K_{c_2}^2}\right) \sum_\mu D_\mu Rs_\mu(\vartheta_2, \xi_0) S'_\mu(\vartheta_2, \eta) = -\frac{j\omega\varepsilon_1}{K_{c_1}^2} \sum_\nu A_\nu \sum_k A_{\nu,k} R'c_k(\vartheta_1, \xi_0) \times$$
$$\times C_k(\vartheta_1, \eta) + \frac{j\omega\varepsilon_2}{K_{c_2}^2} \sum_\nu A_\nu \frac{R'c_\nu(\vartheta_2, \xi_0)}{Rc_\nu(\vartheta_2, \xi_0)} \sum_k A_{\nu,k} Rc_k(\vartheta_1, \xi_0) C_k(\vartheta_1, \eta). \tag{14}$$

Expressions (13) and (14) contain respectively only odd and even functions of η. Both equations can be converted into functional identities by expanding the functions $C'_\nu(\theta_2, \eta)$ and $S'_\mu(\theta_2, \eta)$ in powers of $S_k(\theta_1, \eta)$ and $C_k(\theta_1, \eta)$, namely

$$C'_\nu(\vartheta_2, \eta) = \sum_k A'_{\nu,k} S_k(\vartheta_1, \eta); \quad S'_\mu(\vartheta_2, \eta) = \sum_k D'_{\mu,k} C_k(\vartheta_1, \eta). \tag{15}$$

3. DISPERSION EQUATIONS FOR THE FUNDAMENTAL WAVES

Taking (15) into account, conditions (13) and (14) each break up into several equalities for the coefficients in the corresponding forms of functions: $S_1(\theta_1, \eta)$, $S_3(\theta_1, \eta)$, $S_5(\theta_1, \eta)$, etc. Each equality contains ν coefficients A_ν and μ coefficients D_μ (a total of μ and ν). For the system of equations to be compatible it is essential that μ and ν be equal to twice the number of terms in the expansion in the parameter k, i.e., to the total number of equations. For practical calculations (bearing in mind the specific calculation conditions $\theta_1 \leq 4$ and $|\theta_2| \leq 4$ specifically for each type of wave) we can take two functions in each sum, since the remaining harmonics are quite small. In this case we obtain the following relations

$$A_1\gamma\left(\frac{1}{K_{c_1}^2} - \frac{1}{K_{c_2}^2}\right) Rc_1(\vartheta_2, \xi_0) A'_{11} + A_3\gamma\left(\frac{1}{K_{c_1}^2} - \frac{1}{K_{c_2}^2}\right) \times$$
$$\times Rc_3(\vartheta_2, \xi_0) A'_{31} + D_1\left[\frac{-j\omega\mu_1}{K_{c_1}^2} Rs_1(\vartheta_2, \xi_0) \frac{R's_1(\vartheta_1, \xi_0)}{Rs_1(\vartheta_1, \xi_0)} + \frac{j\omega\mu_2}{K_{c_2}^2} R's_1(\vartheta_2, \xi_0)\right] \times$$

$$+ D_{11} + D_3\left[\frac{-j\omega\mu_1}{K^2_{c_1}} Rs_3(\vartheta_2, \xi_0)\frac{R's_1(\vartheta_1, \xi_0)}{Rs_1(\vartheta_1, \xi_0)} + \frac{j\omega\mu_2}{K^2_{c_2}} R's_3(\vartheta_2, \xi_0)\right] D_{31} = 0, \tag{16}$$

$$A_1\gamma\left(\frac{1}{K^2_{c_1}} - \frac{1}{K^2_{c_2}}\right) Rc_1(\vartheta_2, \xi_0)\, A'_{13} + A_3\gamma\left(\frac{1}{K^2_{c_1}} - \frac{1}{K^2_{c_2}}\right) Rc_3(\vartheta_2, \xi_0)\, A'_{33} +$$
$$+ D_1\left[\frac{-j\omega\mu_1}{K^2_{c_1}} Rs_1(\vartheta_2, \xi_0)\frac{R's_3(\vartheta_1, \xi_0)}{Rs_3(\vartheta_1, \xi_0)} + \frac{j\omega\mu_2}{K^2_{c_2}} R's_1(\vartheta_2, \xi_0)\right] D_{13} +$$
$$+ D_3\left[\frac{-j\omega\mu_1}{K^2_{c_1}} Rs_3(\vartheta_2, \xi_0)\frac{R's_3(\vartheta_1, \xi_0)}{Rs_3(\vartheta_1, \xi_0)} + \frac{j\omega\mu_2}{K^2_{c_2}} R's_3(\vartheta_2, \xi_0)\right] D_{33} = 0, \tag{17}$$

$$A_1\left[\frac{j\omega\varepsilon_1}{K^2_{c_1}} Rc_1(\vartheta_2, \xi_0)\frac{R'c_1(\vartheta_1, \xi_0)}{Rc_1(\vartheta_1, \xi_0)} - \frac{j\omega\varepsilon_2}{K^2_{c_2}} R'c_1(\vartheta_2, \xi_0)\right] A_{11} +$$
$$A_3\left[\frac{j\omega\varepsilon_1}{K^2_{c_1}} Rc_3(\vartheta_2, \xi_0)\frac{R'c_1(\vartheta_1, \xi_0)}{Rc_1(\vartheta_1, \xi_0)} - \frac{j\omega\varepsilon_2}{K^2_{c_2}} R'c_3(\vartheta_2, \xi_0)\right] A_{31} +$$
$$+ D_1\gamma\left(\frac{1}{K^2_{c_1}} - \frac{1}{K^2_{c_2}}\right) Rs_1(\vartheta_2, \xi_0)\, D'_{11} + D_3\gamma\left(\frac{1}{K^2_{c_1}} - \frac{1}{K^2_{c_2}}\right) Rs_3(\vartheta_2, \xi_0)\, D'_{31} = 0, \tag{18}$$

$$A_1\left[\frac{j\omega\varepsilon_1}{K^2_{c_1}} Rc_1(\vartheta_2, \xi_0)\frac{R'c_3(\vartheta_1, \xi_0)}{Rc_3(\vartheta_1, \xi_0)} - \frac{j\omega\varepsilon_2}{K^2_{c_2}} R'c_1(\vartheta_2, \xi_0)\right] A_{13} +$$
$$+ A_3\left[\frac{j\omega\varepsilon_1}{K^2_{c_1}} Rc_3(\vartheta_2, \xi_0)\frac{R'c_3(\vartheta_1, \xi_0)}{Rc_3(\vartheta_1, \xi_0)} - \frac{j\omega\varepsilon_2}{K^2_{c_2}} R'c_3(\vartheta_2, \xi_0)\right] A_{33} +$$
$$+ D_1\gamma\left(\frac{1}{K^2_{c_1}} - \frac{1}{K^2_{c_2}}\right) Rs_1(\vartheta_2, \xi_0)\, D'_{13} + D_3\gamma\left(\frac{1}{K^2_{c_1}} - \frac{1}{K^2_{c_2}}\right) Rs_3(\vartheta_2, \xi_0) D'_{33} = 0. \tag{19}$$

In expressions (16) — (19) the coefficients $A_{\nu,k}$ and $D_{\mu,k}$ are coefficients of the expansion of the functions $C_\nu(\theta_2, \eta)$ and $S_\mu(\theta_2, \eta)$ in functions of the parameters θ_1, without introducing additional factors (as was done before).

The multipliers of A_1; A_3; D_1; D_3 in (16) — (19) will be denoted respectively a_{11}, a_{12}, a_{13}, a_{14}; a_{31}, a_{32}, a_{33}, a_{34}; d_{11}, d_{12}, d_{13}, d_{14}, d_{31}, d_{32}, d_{33}, d_{34}. Then the dispersion equation corresponding to (16) — (19) is written in the form

$$\begin{vmatrix} a_{11} & a_{31} & d_{11} & d_{31} \\ a_{12} & a_{32} & d_{12} & d_{32} \\ a_{13} & a_{33} & d_{13} & d_{33} \\ a_{14} & a_{34} & d_{14} & d_{34} \end{vmatrix} = 0. \tag{20}$$

Even in the form (20) the dispersion equation is still too cumbersome for calculations, and further simplifications are desirable. For this purpose we note that in the fundamental mode the most strongly pronounced is the fundamental harmonic of the angular function, i.e., $A_1 \gg A_3$ and $D_1 \gg D_3$. If we also take it into account that $A'_{11} \gg A'_{31}$, $D'_{11} \gg D_{31}$, $A_{11} \gg A_{31}$ and $D_{11} \gg D_{31}$, we see that relations (6) and (18) can be simplified by discarding the small second-order quantities (which amount to about 1 percent of the fundamental terms). As a result we obtain approximate relations which we shall call for brevity the 'fundamental' approximation

$$A_1\gamma\left(\frac{1}{K^2_{c_1}} - \frac{1}{K^2_{c_2}}\right) Rc_1(\vartheta_2, \xi_0)\, A'_{11} + D_1\left[\frac{-j\omega\mu_1}{K^2_{c_1}} Rc_1(\vartheta_2, \xi_0)\frac{R's_1(\vartheta_1, \xi_0)}{Rs_1(\vartheta_1, \xi_0)} +\right.$$
$$\left. + \frac{j\omega\mu_2}{K^2_{c_2}} R's_1(\vartheta_2, \xi_0)\right] D_{11} = 0, \tag{21}$$

$$A_1\left[\frac{j\omega\varepsilon_1}{K^2_{c_1}} Rc_1(\vartheta_2, \xi_0)\frac{R'c_1(\vartheta_1, \xi_0)}{Rs_1(\vartheta_1, \xi_0)} - \frac{j\omega\varepsilon_2}{K^2_{c_2}} R'c_1(\vartheta_2, \xi_0)\right] A_{11} +$$
$$+ D_1\gamma\left(\frac{1}{K^2_{c_1}} - \frac{1}{K^2_{c_2}}\right) Rs_1(\vartheta_2, \xi_0)\, D'_{11} = 0. \tag{22}$$

The system (21) and (22) yields a dispersion equation in the form

$$\gamma^2\left(\frac{1}{K_{c_1}^2}-\frac{1}{K_{c_2}^2}\right)^2\frac{A'_{11}D'_{11}}{A_{11}D_{11}}=-\omega^2\mu_r\varepsilon_r\left[-\frac{\mu_r}{K_{c_1}^2}\frac{R's_1(\vartheta_1,\xi_0)}{Rs_1(\vartheta_1,\xi_0)}+\frac{1}{K_{c_2}^2}\frac{R's_1(\vartheta_2,\xi_0)}{Rs_1(\vartheta_2,\xi_0)}\right]\times$$

$$\times\left[\frac{\varepsilon_r}{K_{c_1}^2}\frac{R'c_1(\vartheta_1,\xi_0)}{Rc_1(\vartheta_1,\xi_0)}-\frac{1}{K_{c_2}^2}\frac{R'c_1(\vartheta_2,\xi_0)}{Rc_1(\vartheta_2,\xi_0)}\right]. \tag{23}$$

Here $\mu_r=\mu_1/\mu_2$; $\epsilon_r=\epsilon_1/\epsilon_2$. In the remaining equations (17) and (19) the products A_1A_{13}, A_3A_{33}, $D_1D'_{13}$, $D_3D'_{33}$ etc. are of the same order of smallness, so that these equations enable us to calculate A_3 and D_3 into terms of A_1 and D_1 and calculate the field structure as well as make other investigations.

Multiplying (23) by $(c^2)^2$, we recast it in a more compact form. We obtain for the odd fundamental wave $(HE)^{o}_{11}$

$$-\frac{\gamma^2}{\omega^2\mu_r\varepsilon_r}\left(\frac{1}{\vartheta_1^2}-\frac{1}{\vartheta_2^2}\right)^2\frac{A'_{11}D'_{11}}{A_{11}D_{11}}=$$

$$=\left[\frac{\varepsilon_r}{\vartheta_1^2}\frac{R'c_1(\vartheta_1,\xi_0)}{Rc_1(\vartheta_1,\xi_0)}-\frac{1}{\vartheta_2^2}\frac{R'c_1(\vartheta_2,\xi_0)}{Rc_1(\vartheta_2,\xi_0)}\right]\left[-\frac{\mu_r}{\vartheta_1^2}\frac{R's_1(\vartheta_1,\xi_0)}{Rs_1(\vartheta_1,\xi_0)}+\frac{1}{\vartheta_2^2}\frac{R's_1(\vartheta_2,\xi_0)}{Rs_1(\vartheta_2,\xi_0)}\right]. \tag{24}$$

The dispersion equation for the even fundamental wave $(HE)^{e}_{11}$ assumes in the case of the fundamental approximation the following form

$$-\frac{\gamma^2}{\omega^2\mu_r\varepsilon_r}\left(\frac{1}{\vartheta_1^2}-\frac{1}{\vartheta_2^2}\right)^2\frac{C'_{11}B'_{11}}{C_{11}B_{11}}=$$

$$=\left[\frac{\varepsilon_r}{\vartheta_1^2}\frac{R's_1(\vartheta_1,\xi_0)}{Rs_1(\vartheta_1,\xi_0)}-\frac{1}{\vartheta_2^2}\frac{R's_1(\vartheta_2,\xi_0)}{Rs_1(\vartheta_2,\xi_0)}\right]\left[-\frac{\mu_r}{\vartheta_1^2}\frac{R'c_1(\vartheta_1,\xi_0)}{Rc_1(\vartheta_1,\xi_0)}+\frac{1}{\vartheta_2^2}\frac{R'c_1(\vartheta_2,\xi_0)}{Rc_1(\vartheta_2,\xi_0)}\right]. \tag{25}$$

4. DISPERSION CHARACTERISTICS OF THE FUNDAMENTAL WAVES

The slowing-down coefficient of the wave in the waveguide will be denoted by $m=\lambda/\lambda_c$. Here λ_c is the wavelength in the system under consideration (in the waveguide) and λ the wavelength in the outer medium at the same frequency. We also use the notation

$$\vartheta_1^2/\vartheta_2^2=\frac{\varepsilon_r\mu_r-m^2}{1-m^2}=\chi,$$

$$\frac{A'_{11}D'_{11}}{A_{11}D_{11}}=\frac{C'_{11}B'_{11}}{C_{11}B_{11}}=D_1,$$

and obtain the dispersion equations in the following form, which is convenient for calculation.

For the $(HE)^{o}_{11}$ wave

$$m^2=-\frac{1}{D_1}\left[\varepsilon_r\frac{1}{1-\chi}\frac{R'c_1(\vartheta_1,\xi_0)}{Rc_1(\vartheta_1,\xi_0)}-\frac{\chi}{1-\chi}\frac{K'c_1(\vartheta_2,\xi_0)}{Kc_1(\vartheta_2,\xi_0)}\right]\times$$

$$\times\left[\mu_r\frac{1}{1-\chi}\frac{R's_1(\vartheta_1,\xi_0)}{Rs_1(\vartheta_1,\xi_0)}-\frac{\chi}{1-\chi}\frac{K's_1(\vartheta_2,\xi_0)}{Ks_1(\vartheta_2,\xi_0)}\right]. \tag{26}$$

For the $(HE)^{e}_{11}$ wave

$$m^2=-\frac{1}{D_1}\left[\varepsilon_r\frac{1}{1-\chi}\frac{R's_1(\vartheta_1,\xi_0)}{Rs_1(\vartheta_1,\xi_0)}-\frac{\chi}{1-\chi}\frac{K's_1(\vartheta_2,\xi_0)}{Ks_1(\vartheta_2,\xi_0)}\right]\times$$

$$\times\left[\mu_r\frac{1}{1-\chi}\frac{R'c_1(\vartheta_1,\xi_0)}{Rc_1(\vartheta_1,\xi_0)}-\frac{\chi}{1-\chi}\frac{K'c_1(\vartheta_2,\xi_0)}{Kc_1(\vartheta_2,\xi_0)}\right], \tag{27}$$

$Rc_1(\theta_1, \xi_0)$ and $Rs_1(\theta_1, \xi_0)$ are as before the even and odd radial Mathieu-Bessel functions of the first order; $Kc_1(\theta_2, \xi_0)$ and $Ks_1(\theta_2, \xi_0)$ are the even and odd radial Mathieu-Macdonald function. By θ_2 we mean here the modulus of the quantity, as is usually done in the notation for the Macdonald functions.

There are no published tables or curves for the radial functions mentioned above, and we therefore had to calculate these functions and their derivatives with respect to ξ especially for this analysis. We used the expansions cited in references [5, 6, and 7].

The calculation of the coefficient D_1 has yielded the following relation:

$$D_1 = -\frac{[B_1^{(1)}(\vartheta_1)B_1^{(1)}(\vartheta_2) - 3B_3^{(1)}(\vartheta_1)B_3^{(1)}(\vartheta_2) + 5B_5^{(1)}(\vartheta_1)B_5^{(1)}(\vartheta_2) - \ldots]}{[A_1^{(1)}(\vartheta_1)B_1^{(1)}(\vartheta_2) - A_3^{(1)}(\vartheta_1)B_3^{(1)}(\vartheta_2) + A_5^{(1)}(\vartheta_1)B_5^{(1)}(\vartheta_2) - \ldots]} \times$$
$$\times \frac{[A_1^{(1)}(\vartheta_1)A_1^{(1)}(\vartheta_2) - 3A_3^{(1)}(\vartheta_1)A_3^{(1)}(\vartheta_2) + 5A_5^{(1)}(\vartheta_1)A_5^{(1)}(\vartheta_2) - \ldots]}{[A_1^{(1)}(\vartheta_2)B_1^{(1)}(\vartheta_1) - A_3^{(1)}(\vartheta_2)B_3^{(1)}(\vartheta_1) + A_5^{(1)}(\vartheta_2)B_5^{(1)}(\vartheta_1) - \ldots]} \qquad (28)$$

Here $A_m^{(n)}$ and $B_m^{(n)}$ are the coefficients of expansion of the even and odd angular Mathieu functions in powers of trigonometric functions $\cos m\eta$ and $\sin m\eta$, which are found in the handbook tables [8].

The result (28) will not change if we interchange θ_1 and θ_2 in the formula. This condition, together with an account of the properties of the series for $B_m^{(n)}(\theta)$ and $A_m^{(n)}(\theta)$ enables us to obtain for $D_1(\theta_1, \theta_2)$ the following form of approximating function

$$D_1 = -1 + a_{22}\vartheta_1^2\vartheta_2^2 + a_{24}\vartheta_1^2\vartheta_2^2(\vartheta_1^2 + \vartheta_2^2) + a_{44}\vartheta_1^4\vartheta_2^4 + \ldots \qquad (29)$$

In (29) we mean by θ_2 the modulus of the real value of θ_2. The calculations yielded for the coefficients of (29) the following values: $a_{22} = 0.404 \cdot 10^{-2}$; $a_{24} = 0$; $a_{44} = 0.44 \cdot 10^{-5}$.

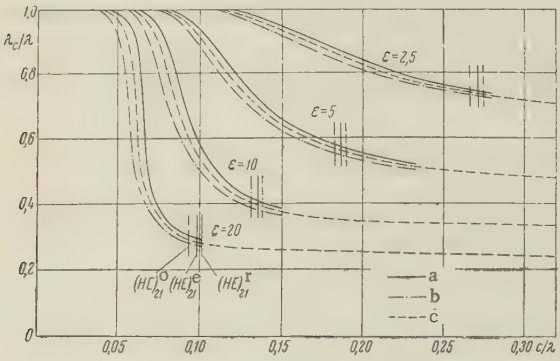


Fig. 2. Dispersion characteristics of the fundamental types of waves:

a — for $(EH)_{11}^{\circ}$ modes of an elliptical waveguide;

b — for $(EH)_{11}^{e}$ modes of an elliptical waveguide;

c — for $(EH)_{11}^{r}$ modes of a round waveguide.

Solutions of the dispersion equations for the case of electric-field polarization parallel to the major axis of the ellipse (odd wave) and transverse to the major axis (even wave), for $\xi_0 = 1$ ($e = c/a = 0.648$), are presented in the form of dispersion characteristics on Fig. 2. For comparison, we show by means of dotted lines the dispersion curves for a dielectric waveguide of round cross section, the horizontal scale being recalculated under the condition that the cross section areas of the waveguides be equal.

It is appropriate to note the wave-process features that are connected with the ellipticity of the cross section. As can be seen from the curves, when the electric field is polarized along the major axis of the ellipse (odd wave) the charcteristic shifts to the left of the curve corresponding to the round cross section. For the even wave the situation is reversed. This means that in the case of the odd and even wave the same concentration of energy in the dielectric occurs at different frequencies. In our case ($\xi_0 = 1$, $e = 0.648$) the difference in frequencies for $\epsilon_r \geq 10$ exceeds 10 percent.

A real wave process can be excited in such a way that the energy is distributed between the even and the odd waves. Owing to the difference in the velocities of wave propagation,

the polarization of the total field will change along the z axis from purely linear to elliptical with a right ellipse, as shown schematically in Fig. 3.

The dispersion curves plotted for different values of the parameter ξ_0 (or b/a) enable us to explain and calculate in many practical cases the effects of a dielectric waveguide produced by rectangular dielectric plates, including hydrodielectrics (ferrites), under different orientations in space or in a metallice waveguide tube.

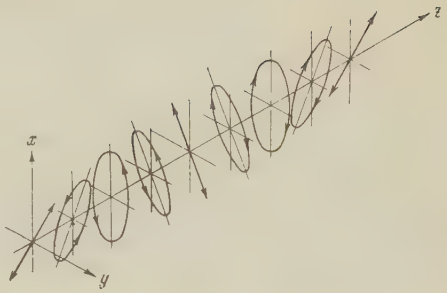


Fig. 3. Pattern of change in polarization of total field along the z axis

5. CRITICAL CONDITIONS FOR HIGHER MODES

The main approximations of the dispersion equation can be obtained also for the case when the angular harmonics of higher order predominate, i.e., for higher modes. The equations of similar to (26) — (28), the only difference being that the lower indices must be replaced by ν.

Near the critical conditions, $-\gamma^2$ approaches $\omega^2\epsilon_2\mu_2$ and $|\theta_2^2| = |c^2K_{c_2}^2| \to 0$. In this connection, we must substitute in the dispersion equations, in place of $K'c_\nu(\theta_2, \xi_2)/Kc_\nu(\theta_2, \xi_0)$ and $K's_\nu(\theta_2, \xi_0)/Ks_\nu(\theta_2, \xi_0)$, the limits of these ratios as $\theta_2 \to 0$.

For practical purpose the greatest interest is attached to an analysis of the dispersion equation in critical conditions for the $(EH)_{21}^{e}$ and $(EH)_{21}^{o}$ modes ($\nu = 2$). We give without the cumbersome intermediate steps the aforementioned asymptotic relations for the case $\nu = 2$:

$$\lim_{\vartheta_2 \to 0} \frac{K'c_2(\vartheta_2, \xi)}{2Kc_2(\vartheta_2, \xi)} = \lim_{\vartheta_2 \to 0} \frac{K's_2(\vartheta_2, \xi)}{2Ks_2(\vartheta_2, \xi)} = -(1 + \vartheta_2^2 a_2), \tag{30}$$

where

$$a_2 = \frac{e^{2\xi} - \frac{1}{3} - e^{-2\xi}}{16}.$$

Using (30) we obtain the dispersion equation under critical conditions for the odd wave in the form

$$\left[-\varepsilon_r\vartheta_2^2 \frac{R'c_2(\vartheta_1, \xi_0)}{Rc_2(\vartheta_1, \xi_0)} + 2\vartheta_1^2(1 + a_2\vartheta_2^2)\right]\left[\vartheta_2^2\mu_r \frac{R's_2(\vartheta_1, \xi_0)}{Rs_2(\vartheta_1, \xi_0)} + 2\vartheta_1^2(1 + a_2\vartheta_2^2)\right] = \pm D_2(\vartheta_1^2 + \mu_r\varepsilon_r\vartheta_2^2)(\vartheta_2^2 + \vartheta_1^2). \tag{31}$$

The analysis yields for the coefficient D_2 the following formula:

$$D_2 = -2^2 \frac{[A_2^{(2)}(\vartheta_1)B_2^{(2)}(\vartheta_2) - 2A_4^{(2)}(\vartheta_1)B_4^{(2)}(\vartheta_2) + 3A_6^{(2)}(\vartheta_1)B_6^{(2)}(\vartheta_2) - \ldots]}{[-2A_0^{(2)}(\vartheta_1)A_0^{(2)}(\vartheta_2) + A_2^{(2)}(\vartheta_1)A_2^{(2)}(\vartheta_2) - A_4^{(2)}(\vartheta_1)A_4^{(2)}(\vartheta_2) + \ldots]} \cdot \frac{[A_2^{(2)}(\vartheta_2)B_2^{(2)}(\vartheta_1) - 2A_4^{(2)}(\vartheta_2)B_4^{(2)}(\vartheta_1) + 3A_6^{(2)}(\vartheta_2)B_6^{(2)}(\vartheta_1) - \ldots]}{[B_2^{(2)}(\vartheta_1)B_2^{(2)}(\vartheta_2) - B_4^{(2)}(\vartheta_1)B_4^{(2)}(\vartheta_2) + \ldots]}.$$

In the operating range of θ_1 and θ_2, the coefficient D_2, like the coefficient D_1 above, can be sufficiently well approximated by a function similar to (29). The calculations yield in this case the following values for the coefficients: $a_{22} = 7.70 \cdot 10^{-3}$; $a_{24} = -2.43 \cdot 10^{-4}$; $a_{44} = 2.60 \cdot 10^{-5}$. The error in the calculation of D_2 by this formula is in effect only in the third significant figure. Using this approximation and confining ourselves to the second power of θ_2, we obtain from (31)

$$\vartheta_1^2 - \frac{1}{2}\vartheta_1^2\vartheta_2^2\left[\varepsilon_r\frac{R'c_2(\vartheta_1,\xi_0)}{Rc_2(\vartheta_1,\xi_0)} + \mu_r\frac{R's_2(\vartheta_1,\xi_0)}{Rs_2(\vartheta_1,\xi_0)}\right] + 2a_2\vartheta_1^4\vartheta_2^2 =$$
$$= \mp\{\vartheta_1^2 + \mu_r\varepsilon_r\vartheta_1^2\vartheta_2^2 + \vartheta_1^2\vartheta_2^2 - a_{22}\vartheta_1^6\vartheta_2^2 - a_{24}\vartheta_1^8\vartheta_2^2\}.$$

If the curly brackets are preceeded by a minus sign and $|\theta\frac{2}{2}| \to 0$ then the equation can be satisfied only when one of the fractions becomes infinite, i.e., one of the denominators vanishes. This corresponds to the root-type critical conditions in the functions Rs and Rc, i.e., to waves of the so-called B branch [1]. In a waveguide with elliptical cross section this branch, as we see, splits in accordance with the even and odd functions. A different sign in front of the brackets yields critical conditions not determined by the roots of the functions Rs and Rc (branch A). For this branch we obtain in the case of the odd mode $(HE)_{21}^{o}$

$$\mu_r\frac{R's_2(\vartheta_1,\xi_0)}{2Rs_2(\vartheta_1,\xi_0)} + \varepsilon_r\frac{R'c_2(\vartheta_1,\xi_0)}{2Rc_2(\vartheta_1,\xi_0)} = -(\mu_r\varepsilon_r + 1) + 2a_2\vartheta_1^2 + (\vartheta_1^4 a_{22} + \vartheta_1^6 a_{24}) \quad (32)$$

and in the case of the even mode $(HE)_{21}^{e}$

$$\mu_r\frac{R'c_2(\vartheta_1,\xi_0)}{2Rc_2(\vartheta_1,\xi_0)} + \varepsilon_r\frac{R's_2(\vartheta_1,\xi_0)}{2Rs_2(\vartheta_1,\xi_0)} = -(\mu_r\varepsilon_r + 1) + 2a_2\vartheta_1^2 + (\vartheta_1^4 a_{22} + \vartheta_1^6 a_{24}). \quad (33)$$

The calculated critical points for the $(HE)_{21}$ modes are shown on the dispersion curves of the fundamental type (see Fig. 2). The points have been calculated by formulas (32) and (33).

In conclusion it must be noted that the expressions obtained for the dispersion equations (26) and (27), for the limiting relations (30), and for the critical conditions (32) and (33) can be converted into the corresponding known formulas for the case of a round dielectric waveguide [1, 2, 3], using the limiting transition: $\xi \to \infty$, $c \to 0$, where ce^{ξ} is a finite quantity. It must be borne in mind here that the differentiation of the radial Mathieu functions that depend on the parameters ξ and θ is carried out everywhere only with respect to the parameter ξ. At the same time, in the theory of round waveguides the radial function is differentiated with respect to the complete argument, which is equivalent to a product of the form θ ch ξ, θ sh ξ or (for large ξ) θe^{ξ}.

CONCLUSIONS

1. Relatively simple approximate expressions are obtained for the dispersion equations of waves in a dielectric waveguide of elliptic cross section, and also for the critical conditions of the nearest higher modes.

2. Formulas have been obtained with which to calculate the field components of the fundamental modes inside and outside the dielectric, with account of the first two spatial harmonics.

3. Dispersion characteristics are obtained for a waveguide with definite ellipticity of the cross section $\xi_0 = 1$.

REFERENCES

1. B.Z. Katsenelenbaum. Propagation of electromagnetic waves along dielectric rods. Sb. nauchnykh trudov, vyp. XI, Izd. Sovetskoye radio, 1948.
2. N.A. Semenov. Parameter of waves in a dielectric waveguide. NDVSh (Radiotekhnika i Elektronika), 1959, 2, 67.
3. G.Z. Aizenberg. Antennas for UHF Waves. Svyaz'izdat, 1957.
4. L.J. Chu. Electromagnetic waves in elliptic hollow pipes of metal. J. Appl. Phys., 1938, 9, 9, 583.
5. M.J.O. Strutt. Lame, Mathieu, and related functions in physics and engineering (in German) Edwards Bros., Ann Arbor, Michigan, 1944.
6. L.J. Chu, J.A. Stratton. Elliptic and spheroidal wave functions. J. Math. and Phys., 1941, 20, 3, 259.

7. N.W. McLachlan. Theory and applications of Mathieu functions. Clarendon Press, Oxford, 1947.
8. J.A. Stratton, P.M. Morse, L.J. Chu, R.A. Hutner. Elliptic, cylinder and spheroidal wave functions. N.Y., 1942.

Received by editor 25 March 1961

EFFECT OF VARIATION OF ELECTRON VELOCITIES ON THE OPERATION OF A BACKWARD WAVE TUBE

V.N. Shevchik and N.I. Sinitsyn

A linear theory for a backward wave tube is developed, with account taken of the variation in the electron velocities and of the attenuation in the system. The results obtained explain the sharp increase in the starting currents of low-voltage backward-wave oscillators on the low end of the electronic tuning range.

INTRODUCTION

Not much research has been done on the influence of variations of electron velocity on the operation of a backward wave tube (BWT). Yet in some cases this problem is of definite interest.

Thus, for example, in a low voltage BWT, where the variation of the electron velocities can be relatively large, a sharp increase is observed in the starting currents at low electron-tuning frequencies [1 — 3].

The variation in electron velocity was first estimated with a one-dimensional approach by Watkins and Rynn [4] and by Johnson [5]. Their analysis describes the phenomena only qualitatively, and the theoretically predicted frequencies at which a helix-type backward-wave oscillator would cease oscillating turned out to be lower than those observed experimentally.

Chang, Shaw, and Watkins [1] investigated this problem by means of a simplified model, in which only the coupling between the slow-wave system and the slow space-charge wave was considered (large values of the space-charge parameter QC). This approach to the solution is rather complicated and lacks clarity.

It is advantageous to estimate the influence of the electron velocity variation on the operation of a backward wave tube by using successive approximations, such as developed by one of the authors [6, 7], since the use of this simple and clear method enables us to account for many effects that are exceedingly difficult to analyze within the framework of the rigorous linear theory [8, 9].

The gist of the method is to calculate the field in the system by successive approximations, the zero-order approximation being the undisturbed state of the system.

Starting with the second approximation, the amplitude of the wave field of the BWT, differs insignificantly from the first approximation, and the field can be represented in the analysis in first approximation, without considerable error, as a perturbation of the 'cold' wave.

The method uses essentially an energy approach, i.e, to determine the wave field we first calculate the interaction power between the electron beam and the field of the 'cold' system.

In the present paper we analyze in the linear approximation the influence of the variation of the electron velocities and the attenutation on the operation of the BWT.

The analysis is based on a procedure developed by Lopukhin [10], with which several problems in microwave electronics can be solved. Using this method, Lopukhin calculated, in particular, the power of interaction between traveling electromagnetic wave and an electron beam, with account of the Coulomb forces and of the electron velocity spread [11]. However, the results of [11] cannot be used in BWT theory because they are valid only for large γl, i.e., for large damping in the system, whereas the method of successive approximations is inaccurate at large values of the damping.

1. DERIVATION OF THE FUNDAMENTAL RELATIONS

In our problem an electron beam, characterized by an electron velocity distribution function having a constant component $f_0(v)$ and an alternating component $f(v, z)e^{j\omega t}$, enters into the half-space $z > 0$, where an electric field of the form

$$E = E_1 e^{\dots} \quad (1)$$

is specified.

The amplitudes of all the variables will be assumed small, and the average electron charge compensated; collisions are neglected. The behavior of the electron beam can then be described by means of the linearized kinetic equations

$$j\omega f + v \frac{\partial f}{\partial z} + \frac{eE_z}{m} \frac{\partial f_0}{\partial v} = 0, \quad (2)$$

where $f_0 \gg |f|$; $|\partial f_0/\partial v| \gg |\partial f/\partial v|$; e and m are the charge and mass of the electron.

The expression for the current density, as usual, is written in the form

$$i_z = e \int_{-\infty}^{\infty} v f(v, z)\, dv.$$

Using the results of integration of (2) with initial conditions $f(v, z)|_{z=0} = f(v, 0)$, we have for an electron beam which is not velocity or density modulated ($f(v, 0) = 0$), and which which enters into the space of interaction with the field of the system [10],

$$i_z = -\frac{j\omega e^2 N_0}{m v_0^2} \int_0^z (u - z) E_z(u) e^{j\beta_e (u-z)} [1 + \varepsilon^2 - \frac{1}{6} \varepsilon^2 \beta_e^2 (u - z)^2 + j\beta_e (u-z)\varepsilon^2 + O(\varepsilon^4)]\, du, \quad (3)$$

where $\epsilon = \Delta v/2v_0 \ll 1$ is the parameter of the electron-velocity spread; $0(\epsilon^4)$ — remainder terms of order ϵ^4 and higher.

The electron velocity variation spread is characterized here by a rectangular distribution function $f_0(v)$, with values

$$f_0(v) = \begin{cases} 0 & \text{for } -\infty < v < v_0 - \frac{\Delta v}{2}, \\ \frac{N_0}{\Delta v} & \text{for } v_0 - \frac{\Delta v}{2} < v < v_0 + \frac{\Delta v}{2}, \\ 0 & \text{for } v_0 + \frac{\Delta v}{2} < v < \infty. \end{cases} \quad (4)$$

where N_0 and v_0 are the mean values of the concentration and electron velocity; Δv — variation of the velocities of the electrons and the beam. The function $f_0(v)$ is expressed with the aid of the delta function $\delta(v)$ in the following manner:

$$j_0(v) = \frac{N_0}{\Delta v}\left\{\int_{-\infty}^{v}\delta\left[v-\left(v_0-\frac{\Delta v}{2}\right)\right]dv-\int_{-\infty}^{v}\delta\left[v-\left(v_0+\frac{\Delta v}{2}\right)\right]dv\right\}. \tag{5}$$

Integrating in (3), we obtain

$$i_z = \frac{j\omega I_0 e E_1}{m v_0^3} e^{-j\beta_e z}\, e^{j\omega t}\, e^{\delta z}\left\{\frac{1+\varepsilon^2}{\delta^2}-\frac{\varepsilon^2\beta_e^2}{\delta^4}-\right.$$
$$-j\,\frac{2\beta_e\varepsilon^2}{\delta^3}+e^{-\delta z}\left[-\frac{1+\varepsilon^2}{\delta^2}(\delta z+1)+\frac{\varepsilon^2\beta_e^2}{6\delta}z^3+\frac{\varepsilon^2\beta_e^2}{2\delta^2}z^2-\right.$$
$$\left.\left.+\frac{\varepsilon^2\beta_e^2}{\delta^3}z+\frac{\varepsilon^2\beta_e^2}{\delta^4}+j\,\frac{3_e\varepsilon^2}{\delta}z^2+j\frac{2\beta_e\varepsilon^2}{\delta^2}z+j\,\frac{2\beta_e\varepsilon^2}{\delta^3}\right]\right\}, \tag{6}$$

where I_0 and V_0 are the current and voltage of the beam, $\delta = \gamma + j\beta_e - j\beta$; $\beta = \omega/v\Phi$; $\beta_e = \omega/v_0$; $v_0 = \sqrt{2e/mV_0}$.

For the case $\epsilon = 0$ we obtain from (6) a previously derived expression for the density of the convection current [6].

The power averaged over one cycle of interaction between the multi-velocity electron beam and the field E_z can be determined from the formula

$$P_e = \frac{1}{2}\int_0^l i_z E_z^*\, dz, \tag{7}$$

where l — length of the system.

Using (5), (6), and (7) we obtain for the electronic interaction power

$$P_e = P_0\,\frac{\xi\varphi_0}{4\Phi_0^3\left(1+\frac{\gamma^2 l^2}{\Phi_0^2}\right)^2}\left\{\frac{e^{2\gamma l}-1}{2\gamma}(1+\xi_1+j\xi_2)\left[2\gamma+j\left(\frac{\gamma^2 l^2}{\Phi_0^2}-1\right)(\beta_e-\beta)\right]+\right.$$
$$(1+\xi_1+j\xi_2)\left[2-\left(2+j\left(1+\frac{\gamma^2 l^2}{\Phi_0^2}\right)\Phi_0\right)e^{\gamma l-j\Phi_0}\right]-$$
$$-\frac{1}{2}\left[\xi_2-j\xi_1+j\varepsilon^2\left(1-\frac{\varphi_0^2}{\Phi_0^2+\gamma^2 l^2}\right)\right]\left[\left(\Phi_0^2\left(1+\frac{\gamma^2 l^2}{\Phi_0^2}\right)\left(\frac{\gamma l}{\Phi_0}+j\right)+\right.\right.$$
$$+2\Phi_0\left(1-\frac{\gamma^2 l^2}{\Phi_0^2}\right)-j4\gamma l+2\,\frac{\left(j+\frac{\gamma l}{\Phi_0}\right)^3}{1+\frac{\gamma^2 l^2}{\Phi_0^2}}\Bigg)e^{\gamma l-j\Phi_0}-2\,\frac{\left(j+\frac{\gamma l}{\Phi_0}\right)^3}{1+\frac{\gamma^2 l^2}{\Phi_0^2}}\Bigg]+$$
$$\left.+j\,\frac{\varepsilon^2\varphi_0^2\Phi_0}{6}\left(1+\frac{\gamma^2 l^2}{\Phi_0^2}\right)e^{\gamma l-j\Phi_0}\right\}, \tag{8}$$

where

$$\xi_1 = \varepsilon^2\left[1-\frac{2\varphi_0}{\Phi_0\left(1+\frac{\gamma^2 l^2}{\Phi_0^2}\right)}-\frac{\varphi_0^2\left(\frac{\gamma^2 l^2}{\Phi_0^2}-1\right)}{\Phi_0^2\left(1+\frac{\gamma^2 l^2}{\Phi_0^2}\right)^2}\right];$$

$$\xi_2 = -\varepsilon^2\left[\frac{2\varphi_0\gamma l}{\Phi_0^2\left(1+\frac{\gamma^2 l^2}{\Phi_0^2}\right)}-\frac{2\varphi_0^2\gamma l}{\Phi_0^3\left(1+\frac{\gamma^2 l^2}{\Phi_0^2}\right)^2}\right];$$

$\Phi_0 = (1 - v_0/v_\Phi)\varphi_0$ — relative transit angle of the electrons moving velocity v_0; $\varphi_0 = \omega l/v_0$; $\xi = V_1/V_0$; $V_1 = E_1 l$.

When $\epsilon = 0$ we obtain from formula (8) the corresponding expressions for the single-velocity approximation [6]. If we are interested in the case of large γl, then by putting $e^{\gamma l} \gg 1$ and retaining only the principle term $\sim e^{\gamma l}$, we arrive at an expression for the power of interaction of the current with the field, obtained in reference [11] (in the absence of Coulomb forces).

2. ANALYSIS OF THE RELATIONS OBTAINED

For the case of a lossless BWT line ($\gamma = 0$), expression (8) can be suitably transformed into the following relations for the active (P_{ea}) and reactive (P_{er}) power components

$$P_{ea} = P_0 \frac{\xi^2\varphi_0}{4\Phi_0^3}\left\{[2(1-\cos\Phi_0) - \Phi_0\sin\Phi_0] + \varepsilon^2\left[2(1-\cos\Phi_0) - \Phi_0\sin\Phi_0 - \right.\right.$$
$$- \frac{2\varphi_0}{\Phi_0}\left(\frac{\Phi_0^2}{2}\cos\Phi_0 + 3(1-\cos\Phi_0) - 2\Phi_0\sin\Phi_0\right) + \frac{\varphi_0^2}{\Phi_0^2}\left(\Phi_0^2\cos\Phi_0 + \right.$$
$$\left.\left.\left. + 4(1-\cos\Phi_0) + \left(\frac{\Phi_0^3}{6} - 3\Phi_0\right)\sin\Phi_0\right)\right]\right\}, \qquad (9)$$

$$P_{er} = P_0 \frac{\xi^2\varphi_0}{4\Phi_0^3}\left\{[2\sin\Phi_0 - \Phi_0(1+\cos\Phi_0)] + \varepsilon^2\left[2\sin\Phi_0 - \Phi_0(1+\cos\Phi_0) + \right.\right.$$
$$+ \frac{2\varphi_0}{\Phi_0}\left(\Phi_0 - 3\sin\Phi_0 + \frac{\Phi_0^2}{2}\sin\Phi_0 + 2\Phi_0\cos\Phi_0\right) - \frac{\varphi_0^2}{\Phi_0^2}\left(\Phi_0 - 4\sin\Phi_0 + \right.$$
$$\left.\left.\left. + \Phi_0^2\sin\Phi_0 - \frac{\Phi_0^3}{6}\cos\Phi_0 + 3\Phi_0\cos\Phi_0\right)\right]\right\}. \qquad (10)$$

Figs. 1 and 2 show plots of the functions $P_{ea}/P_0\xi^2\varphi_0$ and $P_{er}/P_0\xi^2\varphi_0$, obtained from (9) and (10) for different values of the electron-velocity variation parameter ϵ as a function

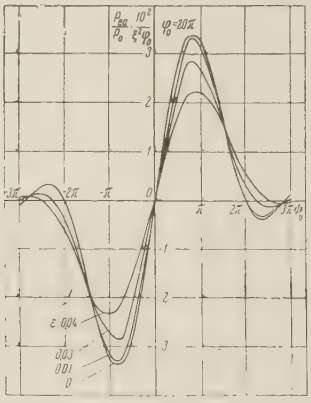


Fig. 1

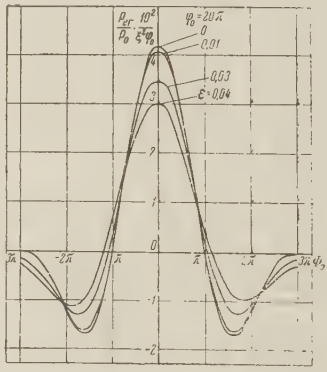


Fig. 2

of Φ_0, the latter changing as the result of changes in the phase velocity of the wave at constant values of l, ω, and v_0. As follows from the curves of Fig. 1, the power of interaction between the electron beam and the field of the system depends strongly on the variation of the electron velocities.

It must be noted that if the electrons are synchronized with the wave ($\Phi_0 = 0$) the multi-velocity beam as a whole draws energy unlike the case $\epsilon = 0$. This result has a clear physical interpretation: the slower electrons, accelerated by the field of the system, stay longer in a

tube with a fixed interaction length l than the faster ones. This explains also the asymmetry of the curves P_{ea} and P_{er} with respect to Φ_0.

This analysis is true for the power of an electron beam interacting with either the forward or the backward waves, i.e., it permits, in particular, to estimate the reduction in the output high-frequency power of an electronic phase shifter, in which low-voltage electron beams are used.

Separating the real and imaginary parts P_{ea} and P_{er} in (8), we find the amplitude of the field of the 'secondary' wave

$$E_2 = E_{2a} + jE_{2r}, \tag{11}$$

induced in the slow-wave line by the electron beam which is bunched in the field of the 'primary' wave, which we have specified as the zeroth approximation.

If K is the coupling resistance of the slow-wave system, we obtain from the energy-conservation law

$$P_{ea} + jP_{er} = \frac{E_1^2}{2\beta^2 K} - \frac{(E_1 e^{\gamma z} + E_2)^2}{2\beta^2 K} + P_\gamma, \tag{12}$$

where the power loss in the line P_γ is given by

$$P_\gamma = \frac{E_1^2 e^{2\gamma z}}{2\beta^2 K} - \frac{E_1^2}{2\beta^2 K}.$$

From (12) we obtain, using (11) the components of the 'secondary' field

$$E_{2a} = \frac{4P_{ea}}{P_0 \xi^2 \varphi_0} (2\pi CN)^3 E_1 e^{-\gamma z}, \tag{13}$$

$$E_{2r} = -\frac{4P_{er}}{P_0 \xi^2 \varphi_0} (2\pi CN)^3 E_1 e^{-\gamma z}, \tag{14}$$

where

$$C^3 = \frac{I_0 K}{4V_0}; \quad N = \frac{\beta_e l}{2\pi}.$$

Relations (1), (13), and (14) enable us to construct readily the distribution of the amplitude of the field in the tube.

Using the condition for backward-wave oscillation

$$E_1 e^{\gamma l} + E_{2a}(l) + jE_{2r}(l) = 0, \tag{15}$$

we obtain in the general form the starting conditions of the BWT

$$CN = \frac{1}{2\pi} \sqrt[3]{\frac{e^{2\gamma l}}{\frac{4P_{ea}}{P_0 \xi^2 \varphi_0}}}, \tag{16}$$

$$P_{er} = 0. \tag{17}$$

When $\gamma = 0$ we can readily obtain by graphical analysis a plot of the starting current of the backward wave oscillator vs. ϵ (Fig. 3).

We note that this analysis is valid for small CN (on the order of 0.5 and below).

It is interesting to estimate the validity of the approximation used in the analysis. In expression (3) we discarded terms of order ϵ^4 and higher. Inclusion of ϵ^4 in the calculation of the current density and the power of interaction between the electron beam and the field of the system results in a much more complicated expression for P_{ea} and P_{er} and thus makes it difficult to determine the starting current. This calculation was made for the particular case $\Phi_0 = 0$. An estimate has shown that the correction obtained by taking this term into

account is very small; we can therefore employ relation (8) with a sufficient degree of approximation. However, extrapolation of (8) to the case of sufficiently large ϵ can lead to principal errors. For real slow-wave system lengths in low-voltage BWT ($\varphi_0 \simeq 40\pi$), the value of ϵ calculated from the resultant formulas should not exceed 0.02 (in Figs. 1 — 3 we assumed for purposes of illustration $\varphi_0 = 20\pi$, and therefore the formulas are valid up to $\epsilon = 0.04$).

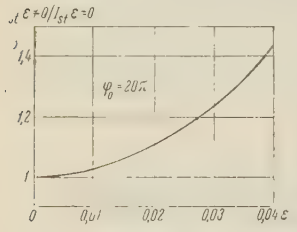


Fig. 3

In our problem we did not state the specific reason for the variation of the electron velocities. In particular cases this can be due to the drooping potential, characteristic of low-voltage BWT. The model considered in this case differs from the specific systems in that the electrons with different velocities are assumed to be mixed together over the entire section of the beam.

In spite of this approximation, we can conclude from our analysis that the reason for the sharp increase in the starting current of the low-voltage backward-wave oscillators towards the low-frequency end of the electron-tuning range is the variation of the electron velocities. Such a conclusion is confirmed by the explanation of the results of [1], in which it was shown experimentally that the reason for the sharp increase in the starting currents is the variation of the electron velocities. On the low-frequency end of the range of electron tuning, a theory that disregards the electron-velocity variation yields a value 12 ma for the starting current, whereas the actual value is 24 ma. The corresponding calculation with the aid of formulas (9) and (10) increases the theoretically obtained starting current to 16.7 ma. On the other hand, if we determine the value of ϵ from the real value of the starting current, the increase is 21.2 ma, which is close to the experimental values. With increasing frequency, the effect due to the electron-velocity spread becomes negligibly small very rapidly.

CONCLUSION

In ordinary backward-wave tubes the effect of the electron-velocity variation is very small and the ordinary single-velocity theory is true. Only on the low-frequency end of the electron-tuning range in low-voltage backward-wave tubes does this variation bring about a sharp increase in the starting currents.

REFERENCES

1. N.C. Chang, A.W. Shaw, D.A. Watkins. The effect of beam cross-sectional velocity variation on backward-wave oscillator current. IRE Trans. Electron Devices, 1959, 6, 4, 437.
2. W.L. Beaver. Backward-wave oscillators for low voltage operation. IRE Convention Record, 1956, 4, 3, 35.
3. N.C. Chang. A helix type backward-wave oscillator in permanent magnet package. Electronic Components Conf. Proc., Los Angeles, California, 1955, p. 47-49.
4. D.A. Watkins, N. Rynn. Effect of velocity distribution of traveling-wave tube gain. J. Appl. Phys., 1954, 25, 11, 1375.
5. H. Johnson. Backward-wave oscillators. Proc. IRE, 1955, 43, 6, 684.
6. V.N. Shevchik. Osnovy elektroniki SVCh [Principles of microwave electronics]. Izd. Sovetskoye radio, 1959.
7. V.N. Shevchik. On the relation between the given-field approximation and the Kompfner successive approximations in the theory of the backward wave oscillator. Radiotekhnika i elektronika, 1959, 4, 1, 147.
8. V.N. Shevchik, D.I. Trubetskov. On the theory of the backward wave tube with periodic focusing of the electron beam. Ibid. 1960, 5, 10, 1734.
9. V.N. Shevchik, Yu.D. Zharkov. Effect of reflections on the operation of a backward wave tube. Ibid, 1960, 5, 12, 2059.
10. V.M. Lopukhin. Excitation of electromagnetic oscillations and waves by means of electron beams. GITTL, 1953.
11. V.M. Lopukhin. Calculation of the power of interaction between a traveling electromagnetic wave and an electron beam with account of Coulomb forces and the electron velocity variation. Izv. vuzov MVO SSSR (Radiofizika), 1958, 1, 2, 27.

Received by editor 29 March 1961

PLASMA CONTROL OF FREQUENCY OF A TRAVELING WAVE OSCILLATOR*

V.D. Ivanova, V.S. Mikhalevskii

An experimental investigation is made of the possibility of controlling the frequency of a traveling wave tube oscillator placed inside a plasma tube by changing the natural frequency of the plasma. A frequency variation which has a jump-like character with a ratio $\sim 1:2$ is obtained in the decimeter band.

The wave resistances and dispersion characteristics are calculated theoretically for the case of an unbounded plasma, as are the dispersion characteristics with account of the finite thickness of the plasma tube. A qualitative agreement between theory and experiment is established.

INTRODUCTION

An analysis of the dispersion properties of a helix placed in an isotropic plasma [1] leads to the conclusion that it is possible to change the dispersion of a slow-wave system of this kind if its geometry is constant. Actually, as shown by experiment [1], the change in the natural frequency ω_0 of the plasma makes it possible to change appreciably the course of the dispersion. In the present paper, as a further development of previous results [1], we report on experimental investigation of the possibility of using the change in dispersion for purposes of controlling the frequency of a TWT oscillator with helical slow-wave system and magnetless focusing of the beam. In this connection, we analyze theoretically the influence of the plasma on the waveguide properties of a single-helix slow-wave line, surrounded on the outside by an isotropic plasma with dielectric constant $\epsilon_2 = 1 - \frac{\omega_0^2}{\omega^2}$; $\omega_0^2 = 4\pi e^2 n_e/m$ where n_e — concentration of the electrons in the plasma; ω — cyclic frequency of oscillations in the system, while e and m are the charge and mass of the electrons. We also determined the equivalent wave resistances K for the case of an unbounded plasma and calculated the dispersion characteristics of a slow-wave system with account of the finite thickness of the plasma layer surrounding the helix.

1. WAVE RESISTANCES

The efficiency of interaction between slow waves propagating in a helix and an electron beam moving along the helix, is characterized by the value of the wave resistance $K = |E_z|^2 / 2\beta^2 P$, where E_z is the electric field component along the helix axis (the z axis); β — phase constant propagation along the slow-wave system; P — total power flux made up of the fluxes inside and outside the helix, defined in a cylindrical system of coordinates as

$$P = \frac{1}{2}\,\mathrm{Re}\iint_S [EH^*]\,dS = \frac{1}{2}\,\mathrm{Re}\iint_S (E_r H_\varphi^* - E_\varphi H_r^*)\,dS;$$

*Reported at the Fourth All-Union Conference on Radio Electronics of the Ministry of Higher and Secondary Education, USSR, Khar'kov, October, 1960.

S is the area of the circle of radius r enclosing the helix. The asterisks denote complex-conjugate values of the electric and magnetic field components.

Performing the calculation of P and K after Gvozdover [2], we obtain by approximating the helix of radius a with an equivalent perfectly conducting cylinder, the following value for K on the helix axis when r = 0

$$K = \frac{R_0}{\pi a^4 \beta^3 k} \frac{1}{\Delta}, \tag{1}$$

where

$$R_0 = \sqrt{\mu_0/\varepsilon_0}; \quad \Delta = \left[\frac{S(\varkappa) I_1(\varkappa)}{\varkappa^2 I_0(\varkappa)} F + \frac{S'(y) K_1(y)}{y^2 K_0(y)} F'\right] I_0^2(\varkappa);$$

$$F = 1 + \frac{\varkappa}{y} \frac{I_1(\varkappa) K_0(y)}{I_0(\varkappa) K_1(y)} A(y, \varkappa, \varepsilon);$$

$$F' = \varepsilon + \frac{y}{\varkappa} \frac{I_1(\varkappa) K_0(y)}{I_0(\varkappa) K_1(y)} A(y, \varkappa, \varepsilon);$$

S and S' are functions of the arguments $\varkappa = a\gamma$ and $y = a\gamma$, in the form

$$S(\varkappa) = \frac{I_1(\varkappa)}{I_0(\varkappa)} - \frac{I_0(\varkappa)}{I_1(\varkappa)} + \frac{2}{\varkappa};$$

$$S'(\varkappa) = \frac{K_0(\varkappa)}{K_1(\varkappa)} - \frac{K_1(\varkappa)}{K_0(\varkappa)} + \frac{2}{\varkappa};$$

$$A(y, \varkappa, \varepsilon) = \frac{\frac{y}{\varkappa} + \varepsilon \frac{I_0(\varkappa) K_1(y)}{I_1(\varkappa) K_0(y)}}{\frac{y}{\varkappa} + \frac{I_0(\varkappa) K_1(y)}{I_1(\varkappa) K_0(y)}};$$

I_0, I_1, K_0, and K_1 — modified Bessel functions. The parameters γ_1 and γ refer to regions with and without plasma, respectively, $\gamma_1 = \sqrt{\beta^2 - \epsilon k^2}$, $\gamma = \sqrt{\beta^2 - k^2}$, $k = \omega/c$, $\beta = \omega/v (\beta^2 > \epsilon k^2)$, where v — phase velocity of the wave propagating along the slow-wave system. The values of the dielectric constant and the permeability, ϵ_1, μ_1 and ϵ_0, μ_0, correspond to a plasma in vacuum; $\epsilon = \epsilon_1/\epsilon_0$.

Table 1

Helix and frequency indices	Helix radius, a, mm	Helix pitch, h, mm	ω_0, mc
1	4.8	1.3	18750
2	4.8	1.8	6250
3	—	—	3100

Fig. 1 shows the dependence of K (in ohms) on λ — the wave length in free space — calculated by means of Eq. (1) for specific values of the helix geometry and ω_0, using the condition $\mu_1 = \mu_0$. The same figure shows for comparison the values of K_h — wave resistance of a helical slow-wave system in the absence of plasma (dotted). Since K has a large range of values and the scale of Fig. 1 cannot be used for all curves, the variation of K is indicated in several cases by means of a corresponding multiplier. The values of ω_0 in megacycles and the geometry of the helices are indicated in Table 1.

Fig. 2 shows the dispersion characteristics, i.e., the dependence of v/c on λ for helical slow-wave systems corresponding to the data indicated in Fig. 1. The calculations are based on Eq. (1) of reference [1]. The first and second numbers on the curves of Figs. 1 — 3 denote respectively the number of the helix as in Table 1 and the frequency ω_0.

As follows from Fig. 1, a characteristic feature of K is that the values of the wave resistances in the presence of a plasma are several orders of magnitude greater than the corresponding values without plasma. This situation remains true also for small λ. Thus, for example, although for the helix 11 the values of K are small in the wavelength range 1 — 10 cm, nonetheless they are two orders of magnitude greater than K_h in this range. On going over from positive to negative dispersion the power flux decreases and passes through zero at a certain value $\lambda = \lambda_{max}$, where it reverses sign. On the dispersion curves (Fig. 2) the

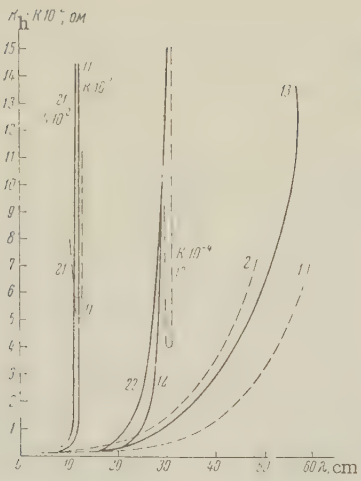


Fig. 1. Wave resistances of helical slow-wave system surrounded by an unbounded layer of plasma.

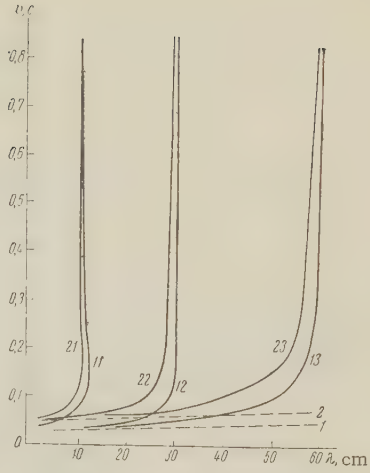


Fig. 2. Dispersion characteristics of helical slow-wave system surrounded by an unbounded layer of plasma.

value $\lambda = \lambda_{max}$ corresponds to a "cutoff" phenomenon defined by the condition that when ω_0 is specified λ_{max} is the upper limit for the slow waves that can propagate in the slow-wave system. Thus, the power flux changes direction at the cutoff point in accordance with the change of direction of the group velocity.

In the negative dispersion region the power flux, whose absolute value passes through a certain maximum, rapidly decreases. The value of K near the cutoff point therefore sharply increases with increasing λ. K then passes through a clearly pronounced minimum. This singularity in the behavior of P and K near λ_{max} is to some degree analogous to the singularity which is characteristic of slow waves in the case of a ribbon type helix, considered in particular in the paper by Sovetov and Sukhov [3].

Table 2

ω_0, mc	ω, mc
18750	16700—16000
6250	6300
3100	3150

With decreasing ω_0, the branches of K corresponding to the portion with negative dispersion (dash-dot curves in Fig. 1) increase and shift towards the larger values of λ. As can be seen from Fig. 1, there is no branch of K of this type for small values of ω_0 (curve 13). With increasing ω_0 (Fig. 2), λ_{max} shifts towards the shorter wave lengths. Calculation shows that cutoff occurs near $\epsilon \approx 0$ and $\omega \approx \omega_0$, i.e., it is determined from the transition from the region of frequencies that are close to but greater than the natural frequency of the plasma towards frequencies smaller than ω_0. A comparison of the values of ω_0, given in Table 2, with the values of ω corresponding to λ_{max} (Fig. 2) confirms this assumption.

2. PLASMA LAYER OF FINITE THICKNESS

In connection with the fact that the plasma layer surrounding the helix has a finite thickness in the actual model of the tube, it is interesting to determine the dispersion properties of the slow-wave system for this case. Using the usual scheme for the solution of problems of this type in a cylindrical coordinate system, and using the approximation of a helically conducting cylinder with perfect conductivity [4], we obtain a dispersion equation in the form

$$\frac{k^2}{\gamma^2}\operatorname{ctg}^2\psi = \frac{y}{\varkappa}\,\frac{\dfrac{I_0(\varkappa)}{I_1(\varkappa)} - \dfrac{y}{\varkappa}\left[\dfrac{BI_0(y) - K_0(y)}{BI_1(y) + K_1(y)}\right]}{\dfrac{y}{\varkappa}\dfrac{I_1(\varkappa)}{I_0(\varkappa)} - \varepsilon\left[\dfrac{B\varepsilon\, I_1(y) \quad K_1(y)}{B\varepsilon I_0(y) - K_0(y)}\right]}, \tag{2}$$

where

$$B = \frac{K_0(w)K_1(g) - \frac{g}{w}K_1(w)K_0(g)}{I_0(w)K_1(g) + \frac{g}{w}I_1(w)K_0(g)};$$

$$B = \frac{\frac{w}{g}K_0(w)K_1(g) - \varepsilon K_1(w)K_0(g)}{\frac{w}{g}I_0(w)K_1(g) + \varepsilon I_1(w)K_0(g)}, \qquad g = b\gamma; w = b\gamma_1;$$

$$\varkappa = a\gamma;, \quad y = a\gamma_1;$$

ψ — angle of rise of the helix; a and b — internal and external radii of the plasma layer surrounding the helix; $\mu_1 = \mu_0$. As $b \to \infty$, B and B_ϵ tend to zero and the dispersion equation (2) goes into relation (1) from [1].

Fig. 3 shows the dispersion characteristics for specific values of the geometry of the helices and ω_0, as determined by Table 1, calculated for different values of b — a. Typical of the dispersion characteristics is the presence of two branches, corresponding to the short and long wave regions. The left-hand short-wave branches remain practically constant as b/a varies over a wide range, and coincide with the dispersion

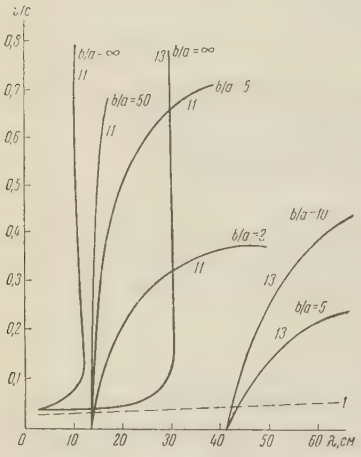


Fig. 3. Dispersion characteristics of helical slow-wave system surrounded by a plasma layer of finite thickness.

characteristics for the case of an unbounded plasma. Actually, only at the large values $\omega_0 \simeq 10^4$ mc does the left branch of the dispersion characteristic for a finite plasma layer shift somewhat to the left of the characteristic when $b/a = \infty$. The right branches become more gently sloping with decreasing b/a, and approach the dispersion characteristic in the absence of plasma. As the phase velocity decreases, the right-hand branches tend to a certain limiting value $\lambda = \lambda_0$, determined by the relation $\omega = \omega_0/\sqrt{2}$. The latter can be obtained from (2) with $y \gg 1$, $w \gg 1$ and $\varkappa \gg 1$, $g \gg 1$. In this case the dispersion equation yields two limiting values of the phase velocity: $v = c \sin \psi$ for short waves and $v = 0$ for $\omega = \omega_0\sqrt{2}$.

Thus in the case of finite thickness of the plasma layer, the dispersion characteristics of a helical slow-wave system has short and long-wave branches and can vary with ω_0 over a wide range, thus providing the possibility of plasma control of the frequency.

3. EXPERIMENTAL PART

To verify experimentally the possibility of plasma frequency control, we assembled a TWT model with centrifugal-electrostatic focusing of the electron beam, operating under continuous evacuation. The slow-wave system of the tube consisted of a helix with radius a = 5 mm and h = 1.5 mm, which was in direct contact with the internal wall of the bulb of the tube. On the outer side there was sealed, coaxially with the glass bulb of the tube, a larger bulb in which plasma was produced by glow discharge in mercury vapor at a pressure on the order of 0.1 mm Hg. Thus, the helix and the plasma were separated only by a thin layer of glass over the entire length of the slow-wave system; b/a = 5.

The mercury-vapor pressure was continuously monitored during the experiment and maintained constant. The change in the natural frequency of the plasma was by changing the current flowing through the plasma. Fig. 4 shows typical parts of the dependence of the frequency f of the oscillations excited in the plasma on the current I flowing through the plasma, with the geometry and operating conditions of the oscillator maintained constant. When I = 0, the oscillations excited had a frequency of 400 mc. As can be seen from Fig. 4, as I is varied, the changes in frequency are in jumps, from one interval of values to the other.

The jump-like character of the tuning is due greatly to the fact that internal feedback is

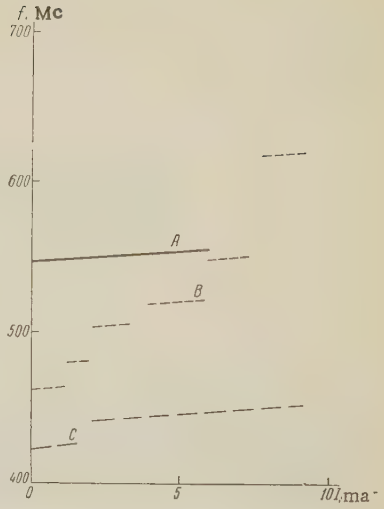


Fig. 4. Dependence of the frequency excited in the TWT on the current flowing through the plasma. A — U_a = 700 v, B — U_a = 780 v, C — U_a = 900 v.

produced in the tube by reflection from the ends of the unmatched helix. The smooth variation in frequency with increasing I, within the limits of each interval, corresponding to excitation near some value of f, has a range of 1 — 2 percent (U_a = 780 v) to 10 percent (U_a = 700 — 900 v). The most effective change in f is observed at a certain fixed value of anode voltage U_a. For this optimum value (U_a = 780 v), the overall change in frequency has a ratio on the order of 1:2. It is easy to see that the variation of f with I, shown in Fig. 4, corresponds to displacement along a certain horizontal line v/c = const, satisfying the condition $v \simeq v_e \sim \sqrt{U_a}$, where v_e — velocity of the electron beam, crossing the dispersion characteristics of Fig. 2 and 3. With this, the excited oscillations shift towards the shorter wave lengths with increasing I and, consequently, with increasing ω_0. Thus, a comparison of the experimental results with those theoretically calculated leads to the conclusion that they are in qualitative agreement and that plasma control of the oscillator frequency is feasible.

REFERENCES

1. V.D. Ivanova, V.S. Mikhalevskii. Dispersion properties of a helix placed in a plasma. Radiotekhnika i elektronika, 1959, 4, 11, 1932.
2. S.D. Gvozdover. Theory of electronic microwave tubes. GITTL, 1956.
3. N.M. Sovetov, A.V. Sukhov. On the calculation of the power flux and coupling resistance in a ribbon type helix. Radiotekhnika i elektronika 1957, 2, 5, 622.
4. V.M. Lopukhin. Excitation of electromagnetic oscillations and waves by means of electron beams. GITTL, 1953.

Received by editor 2 March 1961

BANDWIDTH AND NOISE FIGURE OF TUNNEL-DIODE TUNED AMPLIFIER

L.A. Birger

Exact formulas are derived for the calculation of the bandwidth and the noise figure of a tuned tunnel-diode amplifier. Account is taken of the distributed nature of the resonant-circuit parameters of the losses, and also of the inductance of the leads. The effect of the coupling with the load is analyzed.

The approximate formulas for the bandwidth $2\Delta f$ and the noise figure F of a tuned tunnel-diode amplifier, given in the literature [1, 2, 3], indicate correctly the main features of such amplifiers, but lack certain details useful for a detailed analysis of the amplifier and for an estimate of the validity of the approximate formulas. Thus, for example, the formulas for $2\Delta f$ presume that the tank circuit consists of a lumped capacitance and inductance, whereas high-frequency amplifiers are characterized by distributed constants. Nor is an explicit form given for the dependence of $2\Delta f$, on the operating frequency $\omega_0 = 2\pi f_0$ at a specified gain K_0.

The approximate formulas for F show that it is advantageous to reduce the coupling between the tank circuit and the load, but the limits to which this reduction can be carried are not clear. Finally, the formulas for $2\Delta f$ and F do not take into account the inductances of the leads to the tunnel diode, and while this does not result in noticeable errors at the attained operating frequencies, the limits within which the neglect of inductance is permissible are not indicated.

In order to obtain more accurate expressions for the main characteristics of the tuned amplifier, we carried out the calculations, the main results of which are reported below.

The calculations are based on the universally accepted equivalent circuit of the tunnel diode (Fig. 1a) where R_d is the negative resistance at the operating point; r_d — loss resistance in the diode; C_d — capacitance of the p-n junction; L_d — inductance of the leads. The noise current generator I_1 represents the shot noise in the p-n junction, and generator I_2 represents the thermal noise in the active resistance r_d. The intensities of these noise sources are given by the formulas

$$\overline{I_1^2} = 2eI_0\Delta f, \tag{1}$$

where I_0 — direct current at the operating point, e — electron charge,

$$\overline{I_2^2} = 4kT\Delta f \ \frac{1}{r_d}, \tag{2}$$

where T — temperature of the diode (in degrees Kelvin), and k is Boltzmann's constant.

For convenience in calculation, the network shown in Fig. 1a has been converted to an equivalent "parallel" network (Fig. 1b).

After straightforward but cumbersome calculations we can show that the parameters of the parallel circuit are given by the following formulas

$$g_{de} = -\frac{1}{R_d}\frac{u(\theta)}{w(\theta)}, \tag{3}$$

$$jb_{de} = \frac{j\theta}{R_d} \frac{v(\theta)}{w(\theta)}, \tag{4}$$

$$\overline{I_{1e}^2} = (2eI_0\Delta f)\frac{1}{w(\theta)}, \tag{5}$$

$$\overline{I_{2e}^2} = (4kT\Delta f\ \frac{1}{R_d})\,\xi\,\frac{1+\theta^2}{w(\theta)}, \tag{6}$$

where we put

$$u(\theta) = 1 - \xi - \xi\theta^2; \tag{7}$$

$$v(\theta) = 1 - \frac{1}{\theta_d^2} - \frac{\theta^2}{\theta_d^2}; \tag{8}$$

$$w(\theta) = \theta^2\left(\left(\xi - \frac{1}{\theta_d^2}\right)^2 + \left(1 - \xi - \frac{\theta^2}{\theta_d^2}\right)\right); \tag{9}$$

$$\theta = \omega\tau_d; \tag{10}$$

$$\tau_д = R_d C_d; \tag{11}$$

$$\xi = \frac{r_d}{R_d}; \tag{12}$$

$$\theta_d = \omega_d \tau_d; \tag{13}$$

$$\omega_d = \frac{1}{\sqrt{L_d C_d}}. \tag{14}$$

We recall that the tunnel diode can be used as an amplifier so long as the operating frequency is below a certain limiting value ω_{lim} (or $\theta < \theta_{lim} = \omega_{lim}\tau_d$), at which its equivalent conductance g_{de} vanishes. From (3) and (7) it follows that

$$\theta_{lim} = \omega_{lim}\tau_d = \sqrt{\frac{1-\xi}{\xi}}. \tag{15}$$

Fig. 2 shows the dependence of the admittance g_{de}, jb_{de} on the frequency for the case $\theta_{lim} = 10$.

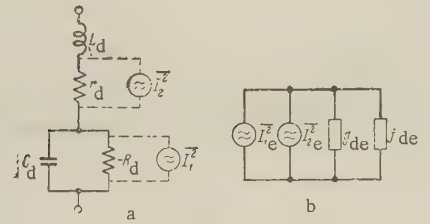


Fig. 1
a — equivalent circuit of tunnel diode;
b — equivalent parallel circuit.

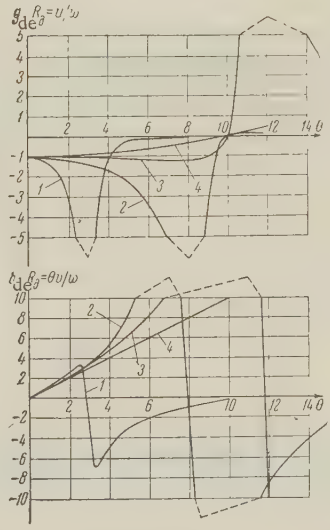


Fig. 2. Dependence of the active and reactive components of the admittance of the tunnel diode on the operating frequency ($\theta_{lim} = 10$):

1 — $\theta_d = 3$; 2 — $\theta_d = 8$; 3 — $\theta_d = 12$; 4 — $\theta_d = \infty$

The character of the functions depends essentially on the relation between the natural resonant frequency θ_d and the limiting frequency θ_{lim}. On approaching the resonant frequency, the absolute value of the conductance increases, and the susceptance first increases, and then decreases and reverses sign.

To ensure stability of the amplifier, the tendency in the construction of tunnel diodes is to reduce the inductance of the leads to an extent such as to satisfy the inequality $\theta_d > \theta_{lim}$.

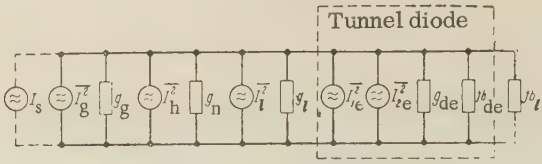


Fig. 3. Equivalent circuit of tuned tunnel-diode amplifier.

Fig. 3 shows the equivalent circuit of a tunnel-diode resonant amplifier, in which g_l — active conductance of the load; g_g — active conductance of the source I_g of the amplified signal; jb_l — susceptance of resonant circuit connected to the tunnel diode, which is tuned at the operating frequency in resonance with the reactance of the diode,

$$b_{de} + b_l = 0 \quad \text{for} \quad \omega = \omega_0. \tag{16}$$

The active losses in the resonance circuit are represented by the conductance g_h, which is best expressed in terms of the ratio η of the "loaded" Q_l to the "unloaded" Q_0 of the resonant circuit:

$$\eta = \frac{Q_l}{Q_0} = \frac{g_h}{g_g + g_l + g_h}; \qquad g_h = \frac{\eta}{1-\eta}(g_g + g_l). \tag{17}$$

The noise generators I_g, I_L and I_h represent the thermal losses of the active resistances:

$$\overline{I_g^2} = 4kT_g \Delta f g_g, \tag{18}$$

$$\overline{I_l^2} = 4kT_l \Delta f g_l, \tag{19}$$

$$\overline{I_h^2} = 4kT \Delta f g_h. \tag{20}$$

Using the circuit of Fig. 3, we can calculate the main characteristics of the tuned amplifier, namely the gain, the bandwidth, and the noise figure.

Let us define the gain (as applied to the conditions under which this parameter is measured) as the square root of the ratio of the power P_L delivered to the load conductance to the maximum power of the signal source $P_S = I_s^2/4g_g$. Using (17) and (3) we get

$$K_0 = \frac{2\sqrt{g_g g_l}}{g_g + g_l g_h + g_{de}} = \frac{2\sqrt{s}}{\frac{1+s}{1-\eta} - \frac{1}{R_d g_g}\frac{u(f)}{w(f)}}, \tag{21}$$

where the parameter

$$s = \frac{g_l}{g_g} \tag{22}$$

characterizes the coupling with the load. We note that $K_0 \to 0$ when $s \to 0$ and $s \to \infty$.

In further calculations we shall assume K_0 specified, in accord with the usual conditions of design of a low-noise amplifier stage.

To calculate the bandwidth

$$2\Delta f = \frac{\omega_0}{2\pi Q}. \tag{23}$$

where Q is the figure of merit in the operating condition ("regenerated" Q), we use the

following general relation [4]:

$$Q = \frac{1}{2}\omega_0 b'_{\Sigma}(\omega_0)\frac{1}{g_{\Sigma}}, \tag{24}$$

where $b'_{\Sigma}(\omega_0)$ is the derivative of the overall susceptance of the circuit with respect to ω

$$b_{\Sigma} = b_{de} + b_{\ell}$$

at the resonant frequency ω_0; and

$$g_{\Sigma} = g_g + g_{\ell} + g_h + g_{de}$$

is the overall conductance.

Formula (24) is valid if the derivative $b'_{\Sigma}(\omega)$ is approximately constant over the pass band. This condition is always satisfied for resonant circuits with a "single hump" resonance characteristic, which will be implied from now on.

Using (24) we can obtain the following formula for the bandwidth

$$2\Delta f = \frac{1-\xi}{2\pi\tau_d K_0}\frac{2}{\sqrt{s}+\sqrt{1/s}}\frac{2}{1+p}\frac{1}{\frac{1}{1-\eta}-\frac{2}{K_0(\sqrt{s}+\sqrt{1/s})}}\Pi(\theta), \tag{25}$$

where we put

$$\Pi(\theta) = \frac{1-\left(\frac{\theta}{\theta_{lim}}\right)^2}{\nu(\theta)\left[1-\frac{2}{1+p}\psi(\theta)\right]}; \tag{26}$$

$$\psi(\theta) = \frac{\theta w'(\theta)}{2w(\theta)} - \frac{\theta v'(\theta)}{2v(\theta)} = \frac{\frac{\theta^2}{\theta_d^2}}{1-\frac{1}{\theta_d^2}-\frac{\theta^2}{\theta_d^2}} + \frac{\theta^2\left(\xi-\frac{1}{\theta_\pi^2}\right)^2 - 2\frac{\theta^2}{\theta_d^2}\left(1-\xi-\frac{\theta^2}{\theta_d^2}\right)}{\theta^2\left(\xi-\frac{1}{\theta_\pi^2}\right)^2+\left(1-\xi-\frac{\theta^2}{\theta_d^2}\right)^2}; \tag{27}$$

$$p = \frac{\omega_0 b'_{\ell}(\omega_0)}{b_{\ell}(\omega_0)}. \tag{28}$$

We note that the parameter p characterizes the degree of distribution of the tank-circuit constants. If the circuit is made up of a lumped inductance, then p = 1. For a circuit made up of short-circuited segment of a long line, p > 1 and amounts to

$$p = \frac{2n\pi}{\sin 2\alpha_0} + \frac{2\alpha_0}{\sin 2\alpha_0}, \tag{29}$$

where n is the number of complete half waves fitting within the length of the circuit; α_0 — "electric" length of the line segment, determined from the condition of resonance with the capacitive susceptance of the tunnel diode

$$b_{de} - \frac{1}{Z_0}\operatorname{ctg}\alpha_0 = 0, \qquad \alpha_0 = \operatorname{arc\,tg}\frac{R_d}{Z_0}\frac{w(\theta)}{\theta v(\theta)}. \tag{30}$$

Here Z_0 is the wave impedance of the line.

Formulas (25), (26), and (27) are rather complicated. However, if we take into consideration the values of the parameters typical of real diodes, these formulas can be greatly simplified. Thus, when

$$\xi \ll 1, \qquad \frac{1}{\theta_d^2} \ll 1, \qquad \frac{\theta_{lim}}{\theta_d^2} \simeq \frac{L_d}{C_d R_d}\frac{1}{\sqrt{R_d r_d}} \ll 1 \tag{31}$$

The function $\pi(\theta)$ (Eq. (26)) assumes the following form

$$\Pi(\theta) \simeq \frac{1-\left(\frac{\theta}{\theta_{\lim}}\right)^2}{1-\left(\frac{\theta}{\theta_d}\right)^2\left(1-\frac{2}{1+p}\right)} \tag{32}$$

If the external resonant circuit of the amplifier is made of a lumped inductance, so that the distribution parameter $p = 1$, we get

$$\Pi(\theta) = 1-\left(\frac{\theta}{\theta_{\lim}}\right)^2. \tag{33}$$

In many cases the frequency of the diode is so high that an inequality stronger than (31) is fulfilled

$$\left(\frac{\theta_{\lim}}{\theta_d}\right)^2 \sim \frac{L_d}{r_d}\frac{1}{\tau_d} \ll 1. \tag{34}$$

Then $\pi(\theta)$ is given by formula (33) even when $p \neq 1$.

Usually we can assume $K_0 \gg 1$, and consequently

$$\frac{2}{K_0(\sqrt{s}+\sqrt{1+s})} \ll \frac{1}{1+\eta}. \tag{35}$$

Taking (31), (33), and (35) into account, formula (25) can be replaced by the simpler formula

$$2\Delta f = \frac{1}{2\pi\tau_d K_0}\frac{2(1+\eta)}{(\sqrt{s}+\sqrt{1+s})}\frac{2}{1+p}\left[1-\left(\frac{\theta}{\theta_{\lim}}\right)^2\right]. \tag{36}$$

If we assume $p = s = 1$, $\eta = 0$, $(\theta/\theta_{\lim})^2 \ll 1$, then by substituting these values in (36) we obtain the well known formula for the approximate estimate of the bandwidth of an amplifier:

$$2\Delta f = \frac{1}{2\pi\tau_d K_0}. \tag{37}$$

Comparing (25), (36), and (37), we notice the following. Formula (37) is extremely approximate and is suitable only for a tentative estimate of the bandwidth at lower operating frequencies. Formula (36) reflects the appreciable influence of many parameters of the amplifiers, which must be taken into account in the design; in particular, it shows the narrowing of the bandwidth on approaching the limiting frequency, highlights the appreciable influence of the distribution parameter p of the resonant circuit and of the parameter s of coupling with the load, and even accounts for the influence of losses in the resonant circuit. Naturally, as the losses are increased with the gain maintained constant (relative to the useful load g_l of the amplifier), the bandwidth decreases. Formula (25) differs from (36) principally that it takes into account the inductance of the diode. However, since condition (34) and all the more (31) are satisfied with sufficient reliability, formula (36) gives quite satisfactory results.

Using the circuit of Fig. 3, we can readily calculate the noise factor of the amplifier, which, in view of the parallel connection of all the noise sources, is equal to

$$F = 1 + \frac{I_l^2}{I_g^2} + \frac{I_h^2}{I_g^2} + \frac{I_{e1}^2}{I_g^2} + \frac{I_{e2}^2}{I_g^2}.$$

By transforming we obtain the following exact formula for F:

$$F = 1 + \frac{T}{T_g}\left[1 + \frac{\eta}{1+\eta}(1+s) + \frac{T_l}{T_g}s\right]$$

$$\frac{1+s}{1-\xi}\left[\frac{1}{1-\eta}-\frac{2}{K_0(\sqrt{s}+\sqrt{1/s})}\right]\frac{1}{1-\left(\frac{\theta}{\theta_{\text{lim}}}\right)^2}\left[\frac{eI_0}{2kT_g}+\xi(1+\theta^2)\right]. \tag{38}$$

We note that the noise figure does not depend on the natural resonant frequency θ_d of the diode (i.e., it is independent of the inductedness of the leads).

If $K_0 \gg 1$, and if furthermore $T = T_l = T_g = T_0 = 293°K$, we obtain from (38) the following formula for the relative noise temperature t_n of the amplifier:

$$t_n = F - 1 = \frac{\eta+s}{1-\eta}+\frac{1+s}{(1-\eta)(1-\xi)}\frac{1}{1-\left(\frac{\theta}{\theta_{\text{lim}}}\right)^2}[0{,}02I_0R_d+\xi(1+\theta)^2], \tag{39}$$

where I_0 is in milliamperes and R_d in ohms.

When $(\theta/\theta_{\text{lim}})^2 \ll 1$, $s = 1$, $\eta = \xi = 0$ we obtain from (39) the approximate formula

$$t_n = 1 + 0{,}04I_0R_d, \tag{40}$$

which, like (37), is applicable for tentative estimates.

As follows from (39), on approaching the limiting frequency the relative noise temperature t_n increases without limit. To decrease t_n it is necessary to strive to a maximal reduction of the relative losses ξ and η in the diode and in the tank circuit, and also to weaken the coupling with the load (i.e., decrease as close). A reduction of s, however, entails a reduction in the bandwidth [see (36)], which has a maximum when $s = 1$. Consequently it is necessary in the design to seek for an optimum value of s as applied to the specific purpose of the amplifier. If the main requirement is the minimum noise temperature, a reduction of s to 0.1 — 0.2 is permissible, thereby reducing the bandwidth by 30 or 40 percent relative to its maximum value (retaining the specified gain), but increases t_n by almost a factor of 2.

In conclusion, I am grateful to V. B. Shteinshleiger for useful discussion of this work.

REFERENCES

1. H.S. Sommers, et al. Tunnel-diode for low-noise amplification. IRE Wescon Convention Record, 1959, part 3, p. 3-9.
2. P. Panfield. Noise-performance of tunnel-diode amplifiers. Proc. IRE, 1960, 48, 8, 1478.
3. E.G. Nielson. Noise-performance of tunnel-diode. Proc. IRE, 1960, 48, 11, 1903.
4. S.I. Evtyanov. Doubly tuned self oscillators with lines. Radiotekhnika, 1950, 5, 4, 17.

Received by editor 22 February 1961

THERMIONIC PROPERTIES OF HEXABORIDES AND OTHER INTERSTITIAL COMPOUNDS

B.A. Trigubenko, B.M. Tsarev

Literature data on the emission constant and work function of interstitial compounds with high melting points show a considerable disparity between results obtained by different authors.

Certain investigations carried out by the authors of the article, under conditions of vacuum 10^{-7} — 10^{-6} mm mercury, indicate that the emission properties can fluctuate in a wide range, depending on the operating conditions of the cathode. One of the reasons for such changes can be the contamination of the surface of the high-melting-point compounds by films of oxides of the metals contained in them.

To obtain reliable results it is necessary to carry out the investigations of thermoelectronic emission of such compounds under conditions of higher vacuum and with surfaces that are fairly clean.

The possibility of using certain substances as materials for thermionic cathodes is determined by two basic factors, high thermionic emission and low rate of evaporation at the operating temperature. Even Lafferty [1] has shown that the emission from lanthanum hexaboride is inferior only to the emission of oxide cathodes at low temperatures, and the rate of its evaporation is a minimum at a given value of emission current density, compared with many other substances with high melting points (Zr, Mo, Nb, W, Ta, Th on W). If we take into account, on the other hand, the strong bonds between the boron atoms in the hexaborides, bonds which cause a high melting point (exceeding 2100°C), their high chemical resistance to the action of oxygen, moisture, and HCL, their high mechanical strength which results in their ability to withstand high electric field intensities at the cathode, as well as their stability to ion bombardment and the ability to operate under conditions of poor vacuum (to 10^{-4} and even 10^{-2} mm Hg), then the prospects of using hexaboride cathodes are subject to no doubt, for neither the relatively high operating temperature, nor the reaction with the core material, prevented with the aid of protective layers of carbides, diborides, or silicides, cannot serve as a serious obstacle to the development of cathodes based on hexaborides.

However, the literature does not contain as yet the necessary or sufficiently reliable data for a comparative estimate of thermionic properties and the speeds of sublimation of hexaborides of rare earth or alkali-earth metals. From the summary of the emission properties of hexaborides, listed in Table I, we see that different authors have obtained widely divergent results, and values of the emission constant A which are either exceedingly low or exceedingly high compared with the theoretical value (A_0 = 120.4 a/cm^{2}deg^2), which clearly does not agree with the metal-like nature of the hexaboride, as confirmed by many of their properties. A similar picture can be seen also in the case of other interstitial compounds, such as diborides, carbides, etc., examples of which are listed in Table 2.

Anomalously high values of the constant A have even resulted in a communication [2], in which the author attempted to attribute these values to the influence of assumed low widths of the forbidden band or low activation energy in similar compounds. Recently, however, new values have been published for the emission constants of uranium carbide [3], namely φ_0 = 2.94 ev and A = 33 a/cm^{2}deg^2, which eliminate UC from the list of substances with large A. In this connection, certain results of investigations carried out by the authors of this article on the thermionic emission of hexaborides of certain rare earth metals (La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Er, Y) are of interest. The investigated hexaborides were

Table 1

Thermionic Properties of Hexaborides

hexaboride	φ_0, ev	A, a/cm^2 deg^2	Reference	hexaboride	φ_0, ev	A, a/cm^2 deg^2	Reference
CaB_6	2,86	2,6	[1]	EuB_6	4,9	1000—5000	[8]
SrB_6	2,67	0,14	[1]	GdB_6	2,05	0,84	[5]
BaB_6	3,45	16,0	[1]	TbB_6	2,99	120,4 (?)	[8]
ScB_6	2,96	4,6	[8]	DyB_6	3,53	25,1	[5]
YB_6	2,2	15,0	[6]	HoB_6	3,42	13,9	[5]
LaB_6	2,66	29	[1]	ErB_6	3,37	9,9	[5]
LaB_6	2,68	73	[5]	YbB_6	3,13	2,5	[5]
CeB_6	2,59	3,6	[1]	LuB_6	3,0	0,36	[5]
CeB_6	2,93	~580	[5]	$(HMM)B_6$	2,64	14,0	[1]
PrB_6	3,46	~300	[5]	$(HMM)B_6$	2,89	~370	[7]
NdB_6	3,97	420	[5]	ThB_6	2,92	0,53	[1]
SmB_6	4,4	—	[8]				

Note: U, MM — Zermischmetall

Table 2

Thermionic Properties of Certain Diborides and Carbides

Type of substance	φ_0, ev	A, a/cm^2 deg^2	Reference	Type of substance	φ_0, ev	A, a/cm^2-deg^2	Reference
TiB_2	3,88	~884	[9]	ZrC	2,18	0,31	[11]
VB_2	3,95	35,5	[9]	UC	4,57	730000	[10]
CrB_2	3,36	48,0	[9]	UC	2,94	33	[3]
ZrB_2	3,67	0,5	[9]	$(ZrC)_{0,8}\cdot(UC)_{0,2}$	4,3	66000	[10]
ZrB_2	4,48	35000	[11]	ThC_2	3,2	~200	[12]
ZrC	3,8	134	[10]	ThC_2	3,5	~550	[11]

deposited on a tantalum wire, first coated with a sintered layer of tantalum powder. The thickness of the layers of tantalum and hexaboride powders were strictly monitored within narrow limits with the aid of a MIM-7 microscope. The cathodes prepared in this manner were tested in cylindrical diodes with triple anodes. The diodes were evacuated with an oil diffusion pump in conjunction with a liquid-nitrogen trap. The cathode temperature was measured during the time of processing in the pump and during the measurements in the sealed diodes by means of a MOP-8 micropyrometer. The vacuum in the diodes at the time of sealing ranged from 10^{-7} to 10^{-6} mm Hg. After sealing, a barium absorber was sputtered in the diodes. The work function φ_0 and the emission constant A were determined by means of Richardson lines through measurements of the emission current density j_e for several values of the temperature within a range from 900 to 1100°C. Similar determinations were made for each cathode after a series of operating cycles at a definite time and a constant temperature ranging from 1130° to 1300°C (see caption to Fig. 2).

During operation and the tests, the cathode temperature and the heat load on the anodes were considerably lower than during the processing with the pump connected, so that the possibility of some appreciable deterioration of the vacuum in the sealed tubes was eliminated. The values of φ_0 and A, obtained at different instants of cathode operation, fit quite well the linear plot $\log A = f(\varphi_0)$, and the aggregate of values obtained for a series of cathodes fit within a certain scatter ellipse, the major axis of which can be regarded as the equation of the mean linear dependence of log A on φ_0 for the tested batch of electrodes.

Analogous results with greater or smaller deviations from linearity were obtained also for other hexaborides of rare-earth metals which we investigated. In view of the multiplicity of factors that can affect the behavior of φ_0 and A, various deviations for linearity are of course possible. Table 3 shows a summary of the limiting values of φ_0 and A, which we obtained for hexaborides of various rare-earth metals for cathodes operating on modes analogous to those listed in the caption of Fig. 1. These data, obtained at a residual-gas pressure on the order of 10^{-7} —10^{-6} mm Hg, indicate that the residual gases have a considerable

influence on the behavior of the hexaborides we investigated, since a similar linear dependence of log A on φ_0 was noted earlier by many authors, principally for materials which were variously contaminated by extraneous substances. Thus, for example, Reiman ([4], page 113, Fig. 18) notes the linear relation between the values of log A and φ_0, obtained by Dushman and Evans, and also by Kingdon, for thoriated tungsten at different degrees of coating of its surface with thorium. This dependence was noted also in references [14] and [15] for tungsten and platinum, respectively, at different stages of their purification by calcinating to maximum temperatures. The values of φ_0 and A, obtained by Zwikker [16] for zirconium and hafnium at various instants of purification by calcination, also fit well the line $\log A = f(\varphi_0)$ plotted on a semilogarithmic scale. Finally, one must note the theoretical derivation of the linear relationship between A and φ_0 in reference [17], devoted to thermionic properties of film cathodes. All the foregoing facts indicate that either chemically active metals (zirconium, hafnium, tungsten) and even such an inactive metal as platinum, as well as many interstitial compounds with active metallic components (thorium, uranium, zirconium, rare-earth elements) exhibit sensitivity to the action of oxygen, and possibly other components of the residual gases, when tested for thermionic emission in insufficiently high vacuum.

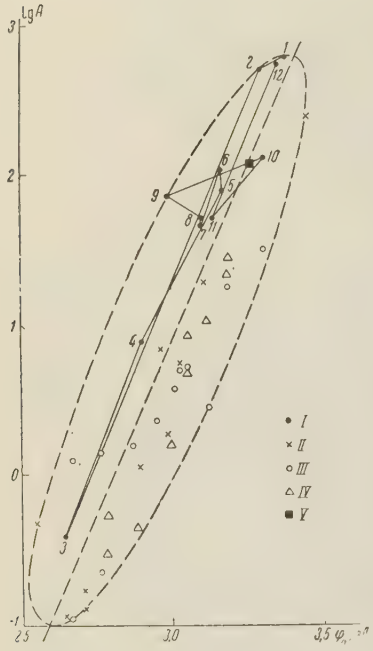


Fig. 1. Variation of the work function φ_0 and the constant A during the operation of TbB_6 cathodes under the following conditions:

Cathode No. 1 (I) (the numbers denote the stages of the operation, after which the values of φ_0 and A were determined, namely: 1 — 5 min at 1130°C, 2 — 15 min at 1130°C, 3 — 30 min at 1130°C, 4 — 60 min at 1130°C, 5 — 120 min at 1130°C, 6 — 120 min at 1150°C, 7 — 120 min at 1175°C, 8 — 120 min at 1200°C, 9 — 120 min at 1225°C, 10 — 120 min at 1250°C, — 11 120 min at 1275°C, 12 — 120 min at 1300°C);
cathode No. 2 (II);
cathode No. 3 (III);
cathode No. 4 (IV);
V — average value of φ_T (with A = 120.4 a/cm^{2}deg^2) for TbB_6, which amounts to 3.26 ev (for T = 1300°K).

It must be noted that changes in the work function, accompanied by changes logarithmically linear in A as a function of φ_0, can be observed also when the emission current

Table 3

Limits of variation of thermionic constants of hexaborides in operation under conditions listed in the caption of Figure 1 (from data obtained for several cathodes produced for each type of boride).

hexaboride	Range of variation of the work function φ_0, ev	Range of variation of the emission constant A, a/cm^{2}deg^2
LaB_6	1,62—3,02	$2\cdot10^{-4}$—312
CeB_6	2,4—5,3	$5\cdot10^{-2}$—$5\cdot10^{8}$
PrB_6	2,8—3,5	0,63—160
NdB_6	1,9—3,3	$5\cdot10^{-4}$—160
SmB_6	2,35—4,1	$6\cdot10^{-5}$—~1600
EuB_6	2,6—3,9	0,01—~100
GdB_6	1,7—3,8	$3\cdot10^{-6}$—~63
TbB_6	2,5—3,5	0,06—~630
ErB_6	2,4—4,1	$3\cdot10^{-4}$—~4000
YB_6	2,3—3,3	0,06—~160

Table 4

Comparison of the values of the work function for hexaborides and oxides of rare-earth metals

hexaboride	φ_T hexaboride (1700°K)	φ_T oxide (1700°K)	oxide	hexaboride	φ_T hexaboride (1700°K)	φ_T oxide (1700°K)	oxide
LaB_6	2,86*,2,87**	3,10	La_2O_3	EuB_6	4,35—4,6*	3,38	Eu_2O_3
CeB_6	3,30*	3,21	Ce_2O_3	GdB_6	2,78*	3,29	Gd_2O_3
PrB_6	3,35*	3,48	Pr_2O_3	TbB_6	3,26***	3,3	Tb_2O_3
NdB_6	3,31*	3,3	Nd_2O_3	ErB_6	3,74*	3,33	Er_2O_3
SmB_6	3,7*	3,21	Sm_2O_3	YB_6	2,52*	3,5	Y_2O_3

* Calculated from data of Table 1.
** Calculated from the equation $\varphi_T = 2.66 + 1.23 \cdot 10^{-4}T$ [18].
*** Determined from the data of Fig. 1 for T = 1300°K.

density is maintained constant at a fixed value of the cathode temperature. In this case, the emission will decrease with increasing φ_0 at the lower values of T, and will increase with φ_0 at higher values of T, as shown in Fig. 2. Therefore merely an investigation of the cathode over a sufficiently wide range of its temperature will disclose the influence of the residual gases and other factors on its thermionic properties.

In addition to direct action of the residual gases on the emission of interstitial compounds, many other factors can also be influential, primarily changes in the phase composition of these compounds and their reactions with the base material.

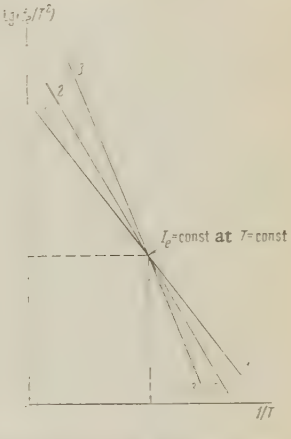


Fig. 2. Possible positions of the Richardson lines (consequently also the variation of φ_0 and A with the emission current density maintained constant at T = const). The relation $\varphi_{01} < \varphi_{02} < \varphi_{03}$ holds for lines 1, 2, and 3.

Changes in the phase composition can be observed when the investigated compound is insufficiently pure, as for example in the case of contamination of the hexaborides by tetra-, tri-, and diborides of the same metal, which can gradually be transformed into hexaborides, which are more stable under high-vacuum conditions, under prolonged high-temperature processing. On the other hand, reaction with the base material, as already noted in [1], leads to a more intense liberation of the metallic component of the boride on the surface of the cathode, and consequently to the formation there of a more stable film of metal oxides under the influence of the residual oxygen. However, in the absence of noticeable reaction with the base, even in the case of pure single-phase composition of the investigated compound, one must apparently expect the surface of the hexaborides to contain a film of oxides of the metal present in the hexaborides, diffusing to the surface and oxidized there by the residual oxygen. Unfortunately, the lack of necessary data on the rate of diffusion of metals in hexaborides and on the evaporation of their oxides from the surface does not enable us to establish any connection whatever between the behavior of φ_0 and A of hexaborides of different metals and the rate of sublimation of their oxides.

To prove the above hypothesis that hexaborides and other interstitial compounds can be contaminated by the oxides of the metals contained in them, we list in Table 4 a summary of the true work functions (φ_T at A = 120.4 a/cm^{2}deg^2) for hexaborides and oxides of certain rare-earth metals, from which we see slight differences in the values of φ_T of the borides and oxides of each of the metals. In addition, a comparison of the behavior of the emission constants φ_0 and A for hexaborides of different rare-earth metals with the data on the rate of evaporation of their oxides, published in [13], shows that high values of the constant A are observed in hexaborides of metals with non-volatile oxides (Y, Gd, Dy, Ho, Er, Lu, and Th).

In view of all the foregoing, reliable determinations of the thermionic properties of hexaborides and other interstitial compounds, particularly with chemically-active metallic

components, can be regarded as possible only if the following conditions are satisfied: 1) the cathodes are processed and tested at ultra high vacuum (not lower than 10^{-8} mm Hg); 2) the investigated cathodes are made of the possibly purest substances with single-phase composition, monitored with the aid of X-ray structural analysis before and after investigation for thermionic emission; 3) eliminate the possibility of the cathode substance reacting with the base metal on which the cathodes are deposited.

Since it is uncertain that oxides and other contaminations can be eliminated from the cathode surface by high-temperature processing alone (evaporation), it is desirable to process the cathode surface prior to measurement of the emission by ion bombardment in an atmosphere of a heavy inert gas (argon, krypton-xenon mixture).

REFERENCES

1. J.M. Lafferty. J. Appl. Phys., 1951, 22, 3, 299.
2. E.A. Kmetko. Phys. Rev., 1959, 116, 4, 895.
3. G.A. Haas, J.T. Jensen. J. Appl. Phys., 1960, 31, 7, 1231.
4. A.L. Reiman. Termionnaya emissiya [Thermionic emission]. GITTL, 1940.
5. G.A. Kudintseva, B.M. Tsarev. Radiotekhnika i elektronika, 1958, 3, 3, 428.
6. G.A. Kudintseva, M.D. Polyakova, G.V. Samsonov, B.M. Tsarer. Fizika metallov i metallovedeniye [Physics of metals and metallography]. 1958, 6, 2, 272.
7. G.A. Kudintseva, V.A. Eppel'baum, B.M. Tsarev. Bor. Tr. konferentsii po khimii bora i yego soyedineniy [Boron, Transactions of Conference on the Chemistry of boron and its compounds]. Goskhimizdat, 1958, pages 112-119.
8. G.V. Samsonov, L.Ya. Markovskii, A.F. Zhigach, M.G. Valyashko. Bor, yego soyedineniya i splavy [Boron, Its Compounds and Alloys]. Izd. AN USSR, 1960, page 432, Table 107.
9. G.A. Kudintseva, B.M. Tsarev, V.A. Eppel'baum. Bor, Tr. konferentsii po khimii bora i yego soyedineniy [Boron, Transactions of Conference on the Chemistry of Boron and Its Compounds]. Goskhimizdat, 1958, pp. 106-111.
10. R.W. Pidd, G.M. Grover, D.J. Roehling, E.W. Salmi, J.D. Farr, N.H. Krikorian, W.G. Witteman. J. Appl. Phys., 1959, 30, 10, 1575.
11. D.L. Goldwater, P.E. Haddad. J. Appl. Phys., 1951, 20, 1, 70.
12. N.D. Morgulis, Yu.P. Korchevoi. Atomnaya energiya [Atomic Energy]. 1960, 9, 1, 49.
13. B.S. Kul'varskaya, R.S. Maslovskaya. Radiotekhnika i elektronika, 1960, 5, 8, 1254.
14. I. Langmuir. Phys. Rev., 1913, 2, 450. (See also A.L. Reiman, Thermoionic emission, GITTL, 1940, page 153, Fig. 34).
15. L.A. Du Bridge. Phys Rev., 1928, 31, 236, 912. (See also A.L. Reiman, Thermoionic emission, GITTL, 1940, page 153, Fig. 34).
16. C. Zwikker. Z. Phys., 1935, 30, 578.
17. A. Gehrts. Z. Phys., 1935, 36, 107-110; Z. f. techn. Phys., 1935, 16, 370-373.
18. V.L. Stout. Proc. 4th Nat. Conf. on Tube Techn., N.Y. Univ. Press, 1959, 178-179.

Received by editor 29 March 1961

OPTIMUM OPERATING POINTS ON THE VOLT-AMPERE CHARACTERISTIC OF A THERMOELECTRIC CONVERTER

I. A. Rezgol'

A theoretical investigation is made of the volt-ampere characteristic of a vacuum thermocouple with compensation of the electron space charge. Equations are derived for the connection between resistance of a "matched" load and the internal resistance of the device under two optimal conditions, corresponding to maximum electric power and to maximum efficiency of conversion of the heat into electricity. A graphic interpretation of the results is given.

INTRODUCTION

The principles of the theory of a vacuum thermocouple with compensation of space charge of the electrons have been developed by Ansel'm [1], who derived the equations for the volt-ampere characteristics of the device and also equations for the thermal processes on the electrodes, for the useful electric power delivered to an external load, and for the efficiency of conversion of heat into electricity. It is known that for specified materials and for electrode temperatures, i.e., for constant saturation emission currents from the emitter and collector, the most important conversion parameters, namely the useful electric power W and the efficiency η, reach a maximum at certain fixed values of the external load resistance (for a matched load). Many papers [2—5] contain a theoretical analysis of "power matching", but either lead to very complicated relationships [4], or contain assumptions that are valid only in certain particular cases [2, 3, 5]. As far as we know, the literature contains no papers concerning the problem of "efficiency matching"; yet this problem is of considerable interest, since the parameters of the device operating at maximum efficiency can differ greatly from the parameters of a device operating at maximum power (particularly at relatively low heat losses due to radiation, as will be shown below, see pages 1911—1912 [of Russian original]).

1. ANALYSIS OF OPTIMUM OPERATING MODE OF THERMOCOUPLE

The present article is devoted to an investigation of the position of the optimal (with respect to power and with respect to efficiency) operating points on the volt-ampere characteristic of a thermoelectric energy converter, in which the space charge of the electrons is fully compensated (in practice such a converter can be realized in the form of a diode filled with low-pressure cesium vapor such that the mean free path of the electrons in the gas is considerably greater than the distance between the electrodes; this instrument, using Dobretsov's terminology [4] is a "vacuum thermocouple with compensation of the electron space charge"). We shall consider an idealized vacuum thermocouple, disregarding in the analysis the ohmic voltage drop and the Joule heat in the leads, the inhomogeneity of the work function over the surface of the electrodes, and the influence of the magnetic fields produced by the generated current on the operation of the instrument. Such a simplification enables us to obtain analytical results that are physically clear.

We shall designate the hot electrode (emitter) by the index 1 and the cold electrode (collector) by the index 2. The fixed parameters will be the work functions φ_1 and φ_2, the electrode temperatures T_1 and T_2, and consequently also the emitter and collector emission conditions in the equations

$$W = IV, \tag{1}$$

$$\eta = \frac{W}{Q_1}, \tag{2}$$

which determine the power delivered to an external load and the efficiency of conversion of heat into electricity. Here I — current generated in the thermocouple; V — voltage on external load; Q_1 — heat power delivered to the emitter from the heat source.

The maximum of equation (1) is determined by the trivial relation

$$-\frac{dV}{dI} = \frac{V}{I} = r, \tag{3}$$

where r — resistance of external load.

Since we have neglected the ohmic voltage drop in the leads, the potential difference U between the emitter and the collector is equal to the external-load voltage with the opposite sign; U = -V. This means that for the idealized thermocouple which we are considering the external and internal volt-ampere characteristics are mere images of each other.

The internal resistance of the thermocouple will be defined as

$$R = \frac{dU}{dI} = -\frac{dV}{dI}. \tag{4}$$

From (3) and (4) it follows that maximum power is delivered to the external load if the load resistance equals the internal resistance of the thermocouple, i.e.,

$$\frac{r}{R} = 1. \tag{5}$$

The relation obtained is similar to the known rule for matched loads, that it must be remembered that in this case the term "internal resistance" is to some extent of formal significance, since we cannot speak of any voltage "drop" inside the converter (U < 0).

Let us proceed to an analysis of the maximum efficiency mode. For this purpose we first consider the specific expressions for the quantity Q_1 in the denominator of (2). As is well known [1], the volt-ampere characteristic of a vacuum thermocouple has in general two regions: when $V < \varphi_1 - \varphi_2$ there a field accelerating the electrons exists between the emitter and the collector (region A); when $V > \varphi_1 - \varphi_2$ the field between the emitter and collector slows down the electrons (region B).

Let us first consider region B of the volt-ampere characteristic. It is convenient here to represent Q_1 as a sum of the useful power W and the heat power Q_2 removed from the collector: $Q_1 = W + Q_2$. The value of Q_2 is:

$$Q_2 = I_1\left(\varphi_2 + \frac{2kT_1}{e}\right) - I_2\left(\varphi_2 + \frac{2kT_2}{e}\right) + \Phi, \tag{6}$$

where k is Boltzmann's constant, e the electron charge, I_1 the electron current flowing from the emitter to the collector, $I_2 = I_{2S}$ is the inverse current from the collector to the emitter, and Φ the heat carried from the emitter to the collector by radiation (generally speaking, this term may include also all the losses not connected with the emitted electrons; for example, losses due to the heat conduction of the gas filling the space between the electrodes, due to heat conduction of the emitter lead, etc.). Equation (6) can be written in a more convenient form

$$Q_2 = I\left(\varphi_2 + \frac{2kT_1}{e}\right) + Q_{OB}, \tag{7}$$

where $Q_{OB} = I_{2S}\frac{2k}{e}(T_1 - T_2) + \Phi$ is the heat carried from the emitter to the collector when the circuit is open and I = 0. We shall later use also a different form of equation (7), namely

$$\varkappa_2 = I + \varkappa_{OB}, \tag{7a}$$

where

$$\varkappa_2 = \frac{Q_2}{\varphi_2 + \frac{2kT_1}{e}}; \quad \varkappa_{OB} = \frac{Q_{OB}}{\varphi_2 + \frac{2kT_1}{e}}.$$

Differentiation (7) with respect to the current, we get

$$\frac{dQ_2}{dI} = \varphi_2 + \frac{2kT_1}{e} = \frac{Q_2 - Q_{OB}}{I}. \tag{8}$$

We next determine the maximum of equation (2), for which we must equate the derivative $d\eta/dI$ to zero. Using (8), we obtain the following condition for the maximum conversion efficiency in region B of the volt-ampere characteristic

$$\frac{r}{R} = \frac{Q_2}{Q_{OB}}. \tag{9}$$

Physically this means that at maximum efficiency the ratio of the load resistance to the internal resistance of the thermocouple is equal to the ratio of the heat removed from the collector in the maximum efficiency load to the heat transported through the instrument in the open-circuit mode.

We now turn to region A of the volt-ampere characteristic. Here we have

$$Q_1 = I_{1s}\left(\varphi_1 + \frac{2kT_1}{e}\right) - I_2\left(\varphi_1 + \frac{2kT_2}{e}\right) + \Phi \tag{10}$$

or

$$Q_1 = I\left(\varphi_1 + \frac{2kT_2}{e}\right) + Q_{OA}. \tag{11}$$

In this case $Q_{OA} = I_{1s}\frac{2k}{e}(T_1 - T_2) + \Phi$ has the physical meaning of the heat carried from the emitter to the collector with the circuit open, but only when $I_{1s} \leq I_{2s}$. If $I_{1s} > I_{2s}$, then the open-circuit mode occurs in region B of the characteristic, and then Q_{OA} must be regarded merely as a convenient analytic expression without direct physical meaning.

The second form of (11) will be

$$\varkappa_1 = I + \varkappa_{OA}, \tag{11a}$$

where

$$\varkappa_1 = \frac{Q_1}{\varphi_1 + \frac{2kT_2}{e}}; \quad \varkappa_{OA} = \frac{Q_{OA}}{\varphi_1 + \frac{2kT_2}{e}}.$$

Further we determine with the aid of the expression

$$\frac{dQ_1}{dI} = \varphi_1 + \frac{2kT_2}{e} = \frac{Q_1 - Q_{OA}}{I} \tag{12}$$

the maximum of (2) and obtain the condition for maximum efficiency in region A of the volt-ampere characteristic

$$\frac{r}{R} = \frac{Q_1}{Q_{OA}}. \tag{13}$$

In other words, at maximum efficiency for region A the ratio of the load resistance to the internal resistance of the thermocouple is equal to the ratio of the heat delivered to the

emitter at maximum efficiency to the heat carried through the device at open circuit (or simply to the value of Q_{OA} when $I_{1s} > I_{2s}$).

Equations (9) and (13) can be unified by representing the maximum efficiency condition in the form

$$\frac{r}{R} = \frac{Q_n}{Q_{OA,B}}. \tag{14}$$

From now on the index n will denote the electrode whose surface has the minimum potential in the thermocouples: For region A this electrode is the emitter, and for region B the collector. The other electrode, whose surface has positive potential, will be denoted by the index p.

In order to transform (5) and (14), which define the optimum modes of the thermocouple operation, to a more convenient form we must determine the ratio r/R.

In region B the current generated in the thermocouple is described by the expression

$$I = I_1 - I_{2s} = I_{1s} \exp\left\{\frac{e[(\varphi_1 - \varphi_2) - V]}{kT_1}\right\} - I_{2s}. \tag{15}$$

Hence

$$R = -\frac{dV}{dI} = \frac{1}{I_1}\frac{kT_1}{e}. \tag{16}$$

We introduce the notation $V_1 = I_1 r$ and $V_2 = I_2 r$ (with $V = V_1 - V_2$). We then obtain for the region B

$$\frac{r}{R} = \frac{eV_1}{kT_1}. \tag{17}$$

In the region A the current is

$$I = I_{1s} - I_2 = I_{1s} - I_{2s} \exp\left\{-\frac{e[(\varphi_1 - \varphi_2) - V]}{kT_2}\right\}. \tag{18}$$

The internal resistance of the thermocouple in region A is

$$R = -\frac{dV}{dI} = \frac{1}{I_2}\frac{kT_2}{e}, \tag{19}$$

and consequently

$$\frac{r}{R} = \frac{eV_2}{kT_2}. \tag{20}$$

Substituting (17) in (20) in (5) and (14) we obtain the maximum conditions for useful power:

for region B

$$\frac{eV_1}{kT_1} = 1, \tag{21}$$

for region A

$$\frac{eV_2}{kT_2} = 1 \tag{22}$$

in general form

$$\frac{eI_p r}{kT_p} = 1 \tag{23}$$

and the condition for maximum conversion efficiency:

for region B

$$\frac{eV_1}{kT_1} = \frac{Q_2}{Q_{OB}}, \tag{24}$$

for region A

$$\frac{eV_2}{kT_2} = \frac{Q_1}{Q_{OA}}, \tag{25}$$

in general form

$$\frac{eV_p}{kT_p} = \frac{Q_n}{Q_{OA,B}} = \frac{\varkappa_n}{\varkappa_{OA,B}}. \tag{26}$$

We note that if we neglect the current due to the backward emission of the collector in region B (if we assume $I_{2S} = 0$), then (21) agrees with the equation derived by Moss [2]:

$$V = \frac{kT_1}{e}, \tag{27}$$

and equation (24) assumes the form

$$V = \frac{kT_1}{e}\frac{Q_2}{Q}. \tag{28}$$

2. GRAPHIC INTERPRETATION OF THE ANALYTICAL RESULTS

It can be easily seen that (23) and (26), which determine the optimum operating conditions of the thermocouple, are transcendental with respect to any of the parameters I and V, which fix the position of the operating point on the external volt-ampere characteristic of the device. This equation contains the variable quantities r, I_p, and $Q_n(\varkappa_n)$, each of which can be given an illustrative geometric interpretation on the plot of I = f(V).

1. The external load resistance r is numerically equal to the cotangent of the angle of inclination of the load line to the abscissa axis (the "load line" is defined as a line drawn through the operating point and through the origin 0 of the volt-ampere characteristic).

2. The current I_p, in accordance with equations (15) and (18), is described by the expression $\pm I_p = I + I_{ns}$, where the plus sign pertains to region B and the minus sign to region A. Consequently, shifting the abscissa axis downward (for the region B) or upward (for the region A) by an amount I_{ns}, we transform the plot I = f(V) into the plot I_p = f(V) with origin at the point O'. In this case the curve of the characteristic for the region A lies below the new abscissa axis (XX axis), so that $I_2 = I_p$ and I flow in opposite directions.

3. According to (7a) and (11a), $\varkappa_n$ depends linearly on the current I. Consequently, by drawing a horizontal axis YY at a distance $\varkappa_{OA,B}$ below the abscissa axis, we obtain from the plot I = f(V) the plot of $\varkappa_n = f(V)$ with origin at the point O".

The geometric relations above enable us to present a clear and simple graphical interpretation of the equations (23) and (26), obtained analytically, using specific volt-ampere characteristics of the thermocouple. For simplicity we consider a case in which both optimum operating points lie either in region B (Fig. 1) or in region A (Fig. 2).

The optimum maximum useful power mode corresponding to equation (23) is shown graphically in the following manner: The load line Ma and the line Mb parallel to the ordinate axis, which are drawn through the optimal working point M, intercept on the XX a segment $ab = kT_p/e$ (since $Mb = I_p$, ctg < Mab = r and $ab = I_p r$).

The optimum conversion efficiency mode, corresponding to equation (26), can be graphically interpreted as follows: the load line Nc drawn through the optimum working point N crosses the XX axis at the point c, through which we draw a line hm parallel to the ordinate axis; the point m where the latter line intersects the YY axis is connected by a line with the point N. The lines hm and Nm intercept on the abscissa axis a segment $fg = kT_p/e$ (this follows from the fact that $fm = \varkappa_{OA,B}$ $hm = \varkappa_n$ and $hN = cd = I_p r$).

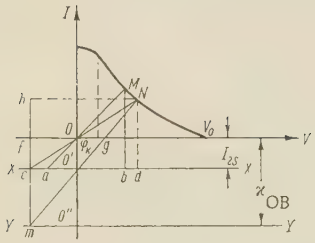


Fig. 1

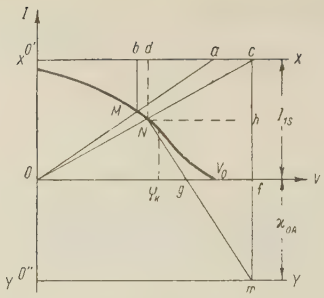


Fig. 2

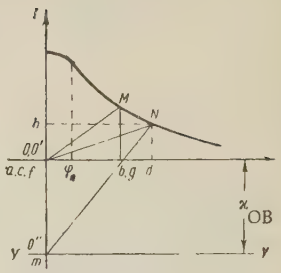


Fig. 3

Fig. 3 shows a plot for the simplified case, corresponding to equations (27) and (28), when the inverse collector emission current is negligibly small ($I_{2S} = 0$) and both optimum operating points lie in region B. Here, unlike the plot of Fig. 1, the XX axis "merges" with the original abscissa axis, and consequently under the graphic constructions indicated above, the points O', a, c, and f coincide with the point O, the point m coincides with the point O'', and the point g coincides with the point b.

Figs. 1—3 show clearly the mutual placement of the two optimum operating points. First, the point N corresponding to the maximum efficiency lies to the right (i.e., in the region of higher values of the external-load voltage), or the point M, corresponding to the maximum useful power. Second, the distance between these points increases with decrease in Q_0, the heat carried from the emitter to the collector when the circuit is open (we recall that the main part of this quantity is the heat transfer due to radiation).

One must point out a possible method of solving (21) — (28) by means of the constructions shown in Figs. 1—3. We can use for the solution a universal plot in dimensionless coordinates, so that it becomes unnecessary to plot specific characteristics for each individual case. We note that in the analyses of references [4] and [5], where the unknown was essentially the operating point corresponding to the region (22), there are given rather complicated methods of solving this problem, of unclear nature, which are either connected with introduction of special tabulated functions [4], or are insufficiently rigorous because certain small quantities are neglected [4, 5]. The graphic solution method is simpler, but we shall not dwell on it in detail, since it can hardly serve as a basis for engineering designs of a thermoelectronic energy converter (in view of the fact that the theoretical model of the vacuum thermocouple is quite idealized).

3. LIMITS OF APPLICABILITY OF THE ANALYTICAL RESULTS

The optimum working points can be situated on the volt-ampere characteristic either in the region B or in the region A (see Figs. 1 and 2). Consequently in each specific case it is necessary to determine the region in which the optimum operating point of interest to us lies. The critical point will be the boundary between the two regions, $V = \varphi_k = \varphi_1 - \varphi_2$. It is easy to show that the optimum operating points are in region B, if the following conditions are satisfied:

for the maximum useful power

$$\frac{e(\varphi_1 - \varphi_2)}{kT_1} \frac{I_{1s}}{I_{1s} - I_{2s}} \leqslant 1. \tag{29}$$

for the maximum conversion efficiency

$$\frac{e(\varphi_1 - \varphi_2)}{kT_1} \frac{I_{1s}}{I_{1s} - I_{2s}} \leqslant \frac{(I_{1s} - I_{2s})\left(\varphi_2 + \frac{2kT_1}{e}\right)}{I_{2s}\frac{2k}{e}(T_1 - T_2) + \Phi} + 1 \tag{30}$$

and that these points are in region A if the following conditions are satisfied:

for the maximum useful power

$$\frac{e(\varphi_1 - \varphi_2)}{kT_2} \frac{I_{2s}}{I_{1s} - I_{2s}} \gtrless 1, \tag{31}$$

for the maximum conversion efficiency

$$\frac{e(\varphi_1 - \varphi_2)}{kT_2} \cdot \frac{I_{2s}}{I_{1s} - I_{2s}} \quad \frac{(I_{1s} - I_{2s})\left(\varphi_1 - \frac{2kT_2}{e}\right)}{I_{1s}\frac{2k}{e}(T_1 - T_2) + \Phi} \gtrless 1. \tag{32}$$

Inequalities (29) and (31) indicate the region in which the maximum useful power is situated, while inequalities (30) and (32) indicate the maximum conversion efficiency mode. One may encounter, however, a case in which both inequalities are not satisfied. Then the equations obtained can no longer be employed, since their roots do not correspond to the real volt-ampere characteristic, and the optimum working point is on the boundary between regions A and B, i.e., the voltage on the matched load will in this case be precisely equal to the contact potential difference between the electrodes. We note also that if all four inequalities (29--32) are not satisfied simultaneously, then we have the only case when the same operating point ($V = \varphi_k = \varphi_1 - \varphi_2$) corresponds simultaneously to the maximum useful power and to the maximum conversion efficiency.

The author is grateful to L.N. Dobretsov for many valuable advices, and also to N.I. Ionov and T.L. Matskevich for useful discussions.

REFERENCES

1. A.I. Ansel'm. Termoelektronnyy vakuumnyy termoelement [Thermionic Vacuum Thermocouple]. Izd. AN SSSR, 1951.
2. H. Moss. Brit. J. Electronics, 1957, 2, 4, 305.
3. G.R. Feaster. J. Electronics and Control. 1958, 5, 2, 142.
4. L.N. Dobretsov. Zh. Tekhn. Fiz. 1960, 30, 4, 365.
5. N.I. Ionov. Zh. Tekhn. Fiz. 1960, 30, 10, 1210.

Received by editor 15 February 1961

CURRENT CONTROL OF THE IGNITION OF A DISCHARGE IN A DEKATRON

F.M. Yablonskiy

An investigation is made of current control of ignition of a gas-discharge gap in a two-pulse dekatron with the aid of an auxiliary discharge. The dependence of the pre-discharge current in the cathode on the current of the auxiliary discharge and on the potential of the cathode relative to the plasma is determined for stationary conditions. The dependence of the ignition voltage under dynamic conditions on the duration in amplitude of the control pulse in the second subcathode, on the rate of growth of the anode voltage, and on the voltage drop on the second subcathode are all investigated.

Current control of ignition, i.e., control of the ignition voltage (U_i) of the main gap by regulating the current in an auxiliary discharge, is widely used in cold-cathode thyratrons.

The operation of the dekatron is also based on reducing the ignition voltage of a discharge produced by current flowing in a neighboring gas-discharge gap.

The thyratron ignition is controlled with an auxiliary discharge produced between an auxiliary anode aa (starting electrode) and the cathode c (Fig. 1a). In the dekatron the auxiliary discharge is between an auxiliary cathode ac (subcathode) and the anode a (Fig. 1b). To produce the auxiliary discharge in the dekatron it is therefore necessary to use a negative voltage pulse, in place of the positive pulse used in the thyratron.

The ignition voltage U_i in the main gap is determined by the current (I_c) flowing in the cathode prior to ignition of the discharge. In the thyratron this current is directly equal to the current of the auxiliary discharge. When the auxiliary discharge is in existence in the subcathode of the dekatron, the cathode potential exceeds the subcathode potential (usually by 40—60 volts). Consequently only part of the auxiliary-discharge current flows to the cathode. To determine I_c in the dekatron we can regard in first approximation the cathode as a negatively-charge electrode — a probe placed in the plasma of the main discharge to the subcathode. The current in such a probe depends on the geometry of the electrodes, on the discharge current to the subcathode, and on the potential of the plasma relative to the probe. The current in such a probe depends on the geometry of the electrodes, on the discharge current flowing in the subcathode, and on the plasma potential relative to the probe.

In a thyratron, when there is no resistance in the cathode circuit, the auxiliary-discharge current does not influence the voltage between the main electrodes. In the dekatron, to the contrary, the auxiliary-discharge current, flowing through R_a, reduces the anode-cathode voltage to a value less than the maintaining voltage (since the subcathode potential is negative with respect to the cathode, and since maintaining voltage of the glow discharge is constant). A discharge can be ignited on the dekatron cathode only when the anode voltage is raised after the discharge at the subcathode is extinguished.

In the present investigation we studied the current control of ignition of a two-pulse dekatron filled with helium and hydrogen, during the transfer of the discharge from the second subcathode to the cathode. We can analyze in similar fashion current control of ignition in the dekatron as the discharge is transferred from the cathode to the first subcathode and from the first subcathode to the second subcathode.

The static characteristics of the dependence of I_c on the voltage between the cathode and the second subcathode (U_{2sc}) with the current maintaining constant in the auxiliary discharge (I_{2sc}) are shown in Fig. 2 (curves 2 and 4). We see from these curves that an increase in U_{2sc} causes a decrease in I_c.

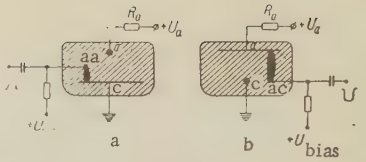


Fig. 1. Control of ignition of the discharge in the thyratron (a) and in a dekatron (b):
a — anode; aa — auxiliary anode; c — cathode; ac — auxiliary cathode (second subcathode)

The potential of the cathode-probe relative to the plasma ($U_{c\,pl}$) is determined by the relation $U_{c\,pl} = U_{pl} - U_{sc}$ (U_{pl} is the potential of the plasma relative to the subcathode). Since $U_{pl} \simeq \Delta U_c$, where ΔU_c is the cathode potential drop, and since in a normal glow discharge $\Delta U_c = \text{const}$, the increment in voltage between the cathode and the subcathode is equal to the decrease in cathode voltage relative to the plasma. Thus, a reduction in the cathode current may be due to a reduction in the potential between the cathode-probe and the plasma.

In a real dekatron circuit, the anode resistance (R_a) is constant, and variation of U_{2sc} influences not only the potential of the cathode-probe relative to the plasma, but also the current in the auxiliary discharge.

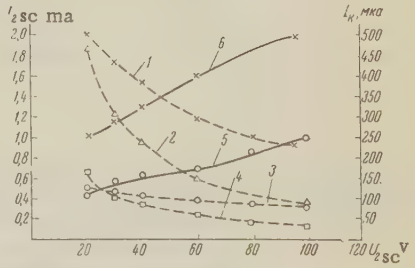


Fig. 2. Static characteristics of the cathode-probe current as a function of the voltage between the cathode and the second subcathode:
1 — $I_c = f(U_{2sc})$, $R_a = 120$ kilohm; 2 — $I_c = f(U_{2sc})$, $I_{2sc} = 1.0$ ma; 3 — $I_c = f(U_{2sc})$, $R_a = 240$ kilohms; 4 — $I_c = f(U_{2sc})$, $I_{2sc} = 0.5$ ma; 5 — $I_{2sc} = f(U_{2sc})$; $R_a = 240$ kilohma; 6 — $I_{2sc} = f(U_{2sc})$, $R_a = 120$ kilohms

It is seen from Fig. 2 that the characteristics $I_c = f(U_{2sc})$, plotted at constant R_a (curves 1 and 3) decrease much more slowly than the characteristics plotted at constant I_{2sc} (curves 2 and 4), this being due to the increase in I_{2sc} with increasing U_{2sc} (see curves 5 and 6).

The foregoing influence of the current of the auxiliary discharge on the voltage on the dekatron anode does not enable us to determine U_i under static conditions, for to increase the anode voltage one must decrease the current in the second subcathode. With slow variation of U_{2sc} this leads to a smooth equalization of the distribution of the currents between the cathode and the second subcathode, so that it is impossible to determine exactly the instant of the discharge on the cathode.

When the dekatron is used in a scalar circuit, the voltage on its subcathodes is applied in pulses, and consequently the increase in the anode voltage to a value U_i is much faster than the rate of change in the distribution of the current between the cathode and the subcathode. Consequently U_i has a definite value, which can be measured with an oscillograph. Oscillograms of the voltage between the anode and cathode and of the current in the second subcathode circuit are shown in Fig. 3, while the principal measurement circuit is shown in Fig. 4. The dekatron anode voltage was measured with an oscilloscope having small input capacitance. To obtain a balanced input, as is required to take oscillograms of the second subcathode current, a pulse transformer (Tr_1) was used. The R_2C_1 network was used to compensate for parasitic surges in the transformer secondary, due to the flow of capacitor currents through the resistance R_1.

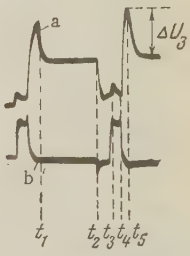


Fig. 3. Oscillograms of voltages and currents in the dekatron: a — voltage between an anode and cathode; b — current in the second subcathode I_{2sc}; U_i — difference between ignition and maintaining voltages in the dekatron.

The oscillogram of Fig. 3a can be explained in the following fashion [1]. During the time $t_1 - t_2$, there exists a cathode discharge, and the anode voltage is equal to the maintaining voltage. At the instant t_2 the discharge is transferred to the first subcathode; the anode potential drops because the first subcathode is negative with respect to the cathode. At the instant t_3 the discharge is transferred to the second subcathode. Since the second subcathode is also negative with respect to the cathode prior to the instant when the control pulse is applied to it, the potential of the anode continues to remain less than the maintaining voltage.

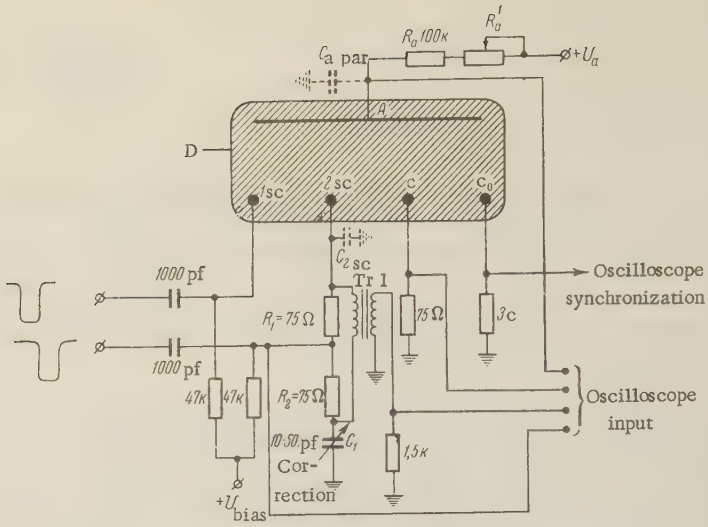


Fig. 4. Circuit for the measurement of the voltages and currents in the dekatron: D — dekatron; $C_{a\ par}$ — parasitic capacitance on the dekatron anode; C_{2sc} — parasitic capacitance in the circuit of the second subcathode; 1sc — first subcathode; 2sc — second subcathode; c — cathode; c_0 — null cathode; Tr1 — pulse transformer.

At the instant t_4 the action of the control pulse on the second subcathode stops, and therefore the anode voltage increases to a value U_i, equal to the voltage necessary to ignite the discharge on the next cathode (instant t_5). For convenience in the measurements we determine with the oscillogram the difference between the ignition voltage and the maintaining voltage (ΔU_i).

From the oscillograms shown in Fig. 3 it is clear that the increase in the voltage on the dekatron anode to a value equal to the ignition voltage U_i is delayed with respect to the instant of cessation of current in the second subcathode circuit, because a considerable time lapse exists between the termination of the current in the second subcathode and the start of the current in the cathode. This feature of the oscillograms is due to the fact that the low internal resistance of the control-pulse source causes a fast change in the voltage (which determines the change in the current in the second subcathode). To the contrary, the increase in the anode voltage with decreasing current is slower, owing to the presence of the parasitic anode capacitance ($C_{a\ par}$), charged through the large resistances R_a and R'_a. As a result, at the instant of termination of the second pulse, the voltage between the anode and the second subcathode becomes lower than the maintaining voltage, and the current in the second subcathode is practically reduced to zero. The increase in voltage between the anode and the cathode, to a value sufficient for ignition of the discharge, is much slower and this causes the delay of the ignition of the discharge at the cathode.

We can thus expect the value of U_i to be determined under dynamic conditions not only by the values of U_{2sc} and I_{2sc}, but also by the rate of increase in the anode voltage (dU_a/dt), a rate that depends on R_a, R'_a, $C_{a\ par}$, the rate of decrease in the voltage on the second subcathode (dU_{2sc}/dt), and also by the duration of the current pulse on the second subcathode (τ).

Fig. 5 shows the dependence of ΔU_i on the amplitude of the pulse in the second subcathode U_{2sc} (measured with respect to the cathode potential), plotted at constant resistance R_a (curves 1 and 2). It is seen from these plots that the increase in U_{2sc} leads to an increase in ΔU_i.

The dependence of ΔU_i on the duration of the control pulse on the second subcathode is shown in Fig. 6 (curves 1 and 3). The plots show that ΔU_i decreases with increasing τ, and this is particularly noticeable at small τ.

Fig. 7 shows the dependences of ΔU_i on the rate of increase on the voltage on the anode dU_a/dt and the rate of decrease on the voltage of the second subcathode dU_{2sc}/dt. The change

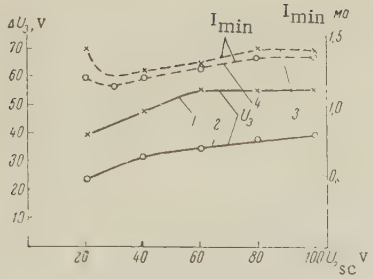


Fig. 5. Dependence of the ignition voltage and of the minimum operating current on the amplitude of the pulse on the second subcathode:
1 — $U_i = f(U_{2sc})$, $\tau = 7$ μsec; 2 — $U_i = f(U_{2sc})$, $\tau = 10$ μsec; 3 — I_a min = $f(U_{2sc})$, $\tau = 7$ μsec; 4 — I_a min = $f(U_{2sc})$, $\tau = 10$ μsec.

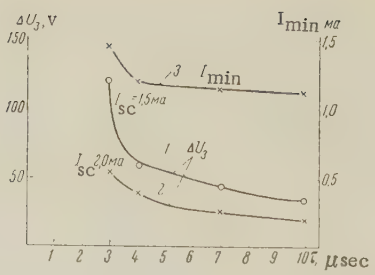


Fig. 6. Dependence of the ignition voltage and of the minimum working current on the duration of the pulse on the second subcathode:
1 — $U_i = f(\tau)$, $I_{2sc} = 1.5$ ma; 2 — $U_i = f(\tau)$, $I_{2sc} = 2.0$ ma; 3 — I_a min = $f(\tau)$.

in the values of dU_a/dt and dU_{2sc}/dt was made by connecting various capacitances to the corresponding electrodes. As can be seen from these curves, ΔU_i decreases with increasing dU_a/dt (curve 1, and increases with increasing dU_{2sc} (curve 2). These relations can also be seen on the oscillograms of Fig. 8, which show the time relationships between the currents in the second subcathode and in the cathode at different values of C_a and C_{2sc}. The oscillogram shown in Fig. 8a corresponds to low parasitic capacitance in the anode circuit,

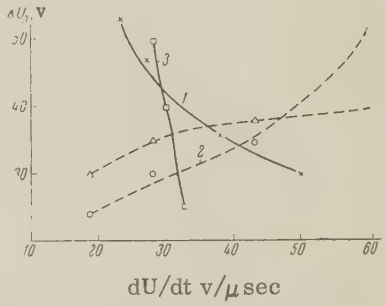


Fig. 7. Dependence of the ignition voltage and minimum working current on the rate of increase in the voltage and on the anode, dU_a/dt, and on the second subcathode dU_{2sc}/dt:
1 — $U_i = f(dU_a/dt)$; 2 — $U_i = f(dU_{2sc}/dt)$; 3 — I_a min = $= f(dU_a/dt)$; 4 — I_a min = $f(dU_{2sc}/dt)$.

i.e., to large dU_a/dt. In this case ΔU_i is small. An increase in C_a (Fig. 8b) results in a decrease in dU_a/dt and in an increase in ΔU_i. The decrease in dU_a/dt can be offset to some extent by increasing C_{2sc}, which reduces dU_{2sc}/dt and ΔU_i (Fig. 8c).

The connection between the rates of decrease and the voltage drop on the second subcathode and the rates of build-up of the voltage on the anode is shown in Fig. 9. The curve of this figure consists of two parts: when $dU/_{2sc}/dt < 28$ v/μsec, the value of dU_a/dt increases with increasing dU_{2sc}/dt; when $dU_{2sc}/dt > 28$ v/μsec, the value of dU_a/dt

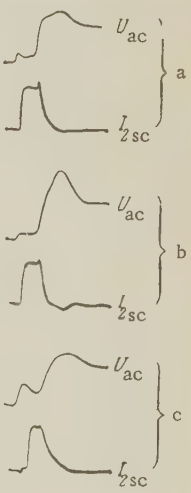


Fig. 8. Oscillograms showing the voltage between the anode and the cathode and the current in the second subcathode for different capacitances in the anode and second subcathode circuits:
a — $C_a \sim 20$ μμf, $C_{2sc} \sim 30$ μμf;
b — $C_a \sim 60$ μμf, $C_{2sc} \sim 30$ μμf;
c — $C_a \sim 60$ μμf, $C_{2sc} \sim 300$ μμf.

remains constant. This character of the curve leads to the conclusion that when dU_{2sc}/dt is small the increase in dU_a/dt is

determined by the rate of decrease of the current on the second subcathode, and when dU_{2sc}/dt is large, it is determined by the discharge of the parasitic capacity into the anode resistance.

The dependence of ΔU_i on the anode current I_a of a dekatron regulated by varying the anode resistance R_a, is shown in Fig. 10 and indicates that an increase in I_a leads to a sharp reduction in ΔU_i.

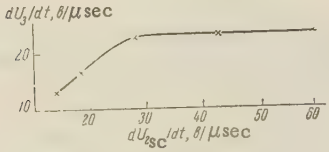


Fig. 9. Dependence of the growth of the anode voltage on the rate of decrease in the voltage on the second subcathode.

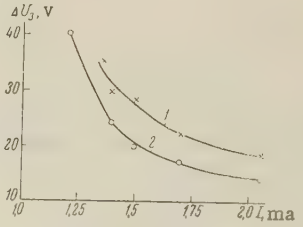


Fig. 10. Dependence of the ignition voltage on the value of the anode current: $1 - \tau = 7$ μsec; $2 - \tau = 10$ μsec.

The foregoing variation of ΔU_i with U_{2sc}, τ, dU_a/dt, dU_{2sc}/dt, and I_a enable us to explain the corresponding variations of the minimum working current $I_{a\ min}$ of the dekatron. The maximum value of ΔU_i under which normal operation of the dekatron (i.e., ignition of a discharge on the cathode, which is adjacent to the arc-carrying second subcathode) is still possible, always corresponds to the minimum working current of the dekatron. Thus, changes in U_{2sc}, τ, dU_a/dt, and dU_{2sc}/dt leading to an increase in ΔU_i are compensated for by a certain increase in $I_{a\ min}$, and conversely, changes in U_{2sc}, τ, dU_a/dt and dU_{2sc}/dt which contribute to a reduction of ΔU_i permit a reduction of $I_{a\ min}$. Figs. 5, 6, and 7 confirm completely the identical qualitative dependence of ΔU_i and $I_{a\ min}$ on U_{2sc}, τ, dU_a/dt and dU_{2sc}/dt. A certain anomaly observed in the course of the dependence $I_{a\ min} = f(U_{2sc})$ at small values of U_{2sc} is apparently connected with disturbances to the normal transfer of the charge from the first subcathode to the second at very small amplitudes U_{2sc}.

The laws derived are also valid for dekatrons filled with inert gas.

The experiments enable us to propose the following qualitative description of the mechanism of charged transfer in the dekatron.

When a negative control pulse is applied to the second subcathode, the main discharge current flows through the subcathode. The cathode operates at that time like a probe surrounded by a unipolar layer, the thickness of which is all the smaller, the greater the ion current flowing in the cathode as a result of the main discharge.

At the instant of termination of the control pulse, the potential of the second subcathode increases, and this causes the current in it to decrease in the potential of the anode (which determines the plasma potential) to increase. Thus, the potential between the cathode-probe and the plasma increases at that instant, and the current of the main discharge decreases. The resultant steady-state distribution of the potential in the unit or layer turns out to be insufficient to screen the field of the cathode; the latter penetrates into the plasma and leads in final analysis to an increase in the thickness of the unipolar layer [2].

The ignition of the cathode discharge can be regarded as a breakdown of the unipolar layer, occurring before the layer is completely rearranged. Inasmuch as pressure and the distance between the electrodes and the dekatron correspond to the right-hand branch of the Paschen curve, the smaller the thickness of the unipolar layer at that instant, the more readily will the breakdown occur.

The thickness of the unipolar layer will decrease with increasing cathode current and vice versa. Thus, the reduction in ΔU_i with decreasing U_{2sc} (Fig. 5) is due to the decrease in the thickness of the ion sheath resulting from the increase of the current I_c flowing in the cathode-probe with decreasing U_{2sc} (Fig. 2). The increase in ΔU_i with decreasing τ (Fig. 6) is due to the fact that the ions from the discharge occurring on the second subcathode

at the instant of time t_3 (Fig. 3) do not have time to diffuse in sufficient quantities towards the cathode-probe, and this results in an increase of the thickness of the unipolar layer around the latter. The reduction in ΔU_i with increasing dU_a/dt and with decreasing dU_{2sc}/dt (Figs. 7 and 8) is due to the reduction in the time lapse between the cessation of the current in the subcathode and the increase in the anode voltage, which causes an increase in the current of the cathode-probe at the instant preceding the breakdown.

Conclusion. An investigation of the dependence of the voltage that ignites the discharge between the anode and the cathode on the duration and amplitude on the control pulse applied to the second subcathode, on the rate of growth of the anode voltage, on the rate of decrease in the voltage on the second subcathode and on the anode current shows that the change in the ignition voltage between the anode and the cathode can be attributed to the change in the thickness of the unipolar layer surrounding the cathode-probe at the instant preceding the breakdown.

The author considers it is his duty to thank V. L. Granovskii for a discussion of the results of this work.

REFERENCES

1. F.M. Yablonskii. Determination of the limiting counting rates of gas-discharge tubes (dekatrons) from the characteristics of the anode-cathode gap. Radiotekhnika i elektronika, 1960, 5, 2, 338.
2. D. Kamke, H.I. Rose. Die Tragerdichte im plasma und ihre Bestimmung mit der Impuls-Sonde. Z. Phys., 1956, 145, 1, 83.

Received by editor 8 February 1961

SOME SPECIAL FEATURES OF THE VOLT-AMPERE CHARACTERISTICS OF NARROW GERMANIUM P-N JUNCTIONS

N.A. Belova, A.N. Kovalev

An investigation was made of the influence of the degree of alloying of germanium on the characteristics of tunnel diodes. It is shown that for narrow germanium p-n junctions, in which the degree of degeneracy in the p-region is greater than the degree of degeneracy in the n-region, the voltage corresponding to the maximum tunnel current is determined by the degree of degeneracy in the p-region.

The density of the tunnel current at the maximum increases with increasing concentration of the majority carriers in the n-region (with constant concentration of the carriers in the p-region) in accordance with the change in the probability of tunneling of the electrons through the potential barrier.

INTRODUCTION

It is shown in [1] that the form of the volt-ampere characteristic of a narrow germanium p-n junction made of n-germanium is greatly influenced by the concentration of the mobile carriers both in the n and in the p regions. In the present article we describe the results of further experiments on the influence of the degree of alloying of germanium of either n or p type on the maximum of the tunnel current.

1. VALUE OF VOLTAGE CORRESPONDING TO MAXIMUM TUNNEL CURRENT

For tunnel diodes made of electronic germanium it has been shown [1] that the voltage corresponding to the maximum tunnel current is practically independent of the concentration of the majority carriers in the n-region and is determined by the degree of alloying of the p-region formed upon addition of the acceptor impurity. In the present paper we show the influence of the concentration of the majority carriers in initial p-type germanium on the form of the volt-ampere characteristics of p-n junctions, i.e., when the concentration of the majority carriers in the p-region is certainly known. For this purpose we made diodes of germanium doped with gallium at a majority-carrier concentration from 1.5×10^{19} to 5×10^{19} cm^{-3}. The region with electron conductivity was obtained by adding to a slab of p-type germanium, indium with arsenic impurity.

For the original germanium the concentration of the mobile p carriers was determined by measuring the Hall constant; the degree of degeneracy μ_p was estimated with the aid of the relation

$$p = \frac{4}{\sqrt{\pi}} \left(\frac{2\pi m_p^* kT}{h^2} \right)^{3/2} \int_0^\infty \frac{\sqrt{x}\, dx}{e^{x-\eta} + 1}, \qquad (1)$$

where m_p^* — effective mass of heavy hole ($0.3\ m_0$); k — Boltzmann's constant; T — absolute temperature; h — Planck's constant; $\eta = \mu_p/kT$; in the calculation we used the tables listed in [2].

The degeneracy calculated in this manner was found to be $\mu_p \sim 3kT$ for $p = 1.5 \times 10^{19}$ cm^{-3}, in the germanium employed.

As indicated above, the n-region was produced by using arsenic, the maximum solubility of which in germanium is 6×10^{19} cm^{-3} [3]. If it is assumed that the alloying produces an n-region with a majority carrier concentration 6×10^{19} cm^{-3}, this corresponds to a degree of degeneracy $\mu_n \simeq 2kT$. Under real conditions, acceptor impurities, namely gallium and indium, also enter into the germanium lattice, along with the arsenic, during the time of formation of the electron region, and therfore the concentration of the majority carriers can in the n region be only less than 6×10^{19} cm^{-3}. Consequently, in the investigated tunnel diodes, made of p germanium, the degeneracy in the hole region is greater than in the electron region, i.e., $\mu_p > \mu_n$.

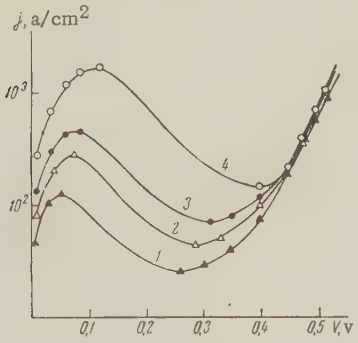


Fig. 1. Volt-ampere characteristics of tunnel diodes made of p-type germanium: The initial germanium has the following: 1 — p = $1.5 \cdot 10^{19}$ cm^{-3}; 2 — p = $2.5 \cdot 10^{19}$ cm^{-3}; 3 — p = $3.5 \cdot 10^{19}$ cm^{-3}; 4 — p = $5.3 \cdot 10^{10}$ cm^{-3}

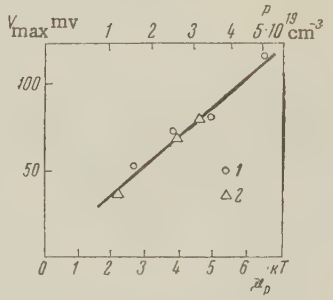


Fig. 2. Dependence of the voltage corresponding to the maximum tunnel current on the degree of degeneracy μ_p (or the concentration p) in the p region. 1 — Diode made of p germanium; 2 — of n germanium

Fig. 1 shows the characteristics of tunnel diodes made of p-type germanium with different majority carrier concentrations. The position of the maximum and the density of the tunnel current are seen to be strongly dependent of the degree of doping of the initial germanium. This agrees with the results of [1] in the sense that the position of the tunnel diode is decisively influenced by the degeneracy in the p region.

Fig. 2 shows the dependence of the voltage corresponding to maximum tunnel current on the degree of degeneracy in the p region. This dependence is linear; the line has a slope 3/5. Thus, for the investigated p-n junctions the following relation holds true

$$eV_{max} \simeq \frac{3}{5}\mu_p. \qquad (2)$$

For comparison we made diodes of n-type germanium with constant concentration in the n region (2×10^{19} cm^{-3}), but with different alloying of the p region. Different concentrations in the p region were obtained by using different contents of gallium in the indium (from 0.5 to 1.5 percent). Since the position of the maximum shifts for these diodes with increasing concentration in the p region, toward the region of larger voltages, one could conclude that $\mu_p > \mu_n$ for these diodes. We therefore used relation (2) to determine the degree of degeneracy in the p region. Fig. 2 shows the calculated values of μ_p for these diodes.

We then compared two pairs of diodes made of n and p germanium with identical values of voltages at the maximum (see Fig. 2). The volt-ampere characteristics of two diodes with V_{max} = 80 mv were perfectly identical. For this pair of p-n junctions we can assume that the concentrations are equal both in the n and in the p regions. The values of the concentrations are listed in the table [p for diodes of n germanium was determined with the aid of (1) and (2)]. On the basis of the satisfactory agreement of the volt-ampere characteristics of these diodes we can conclude that the concentration of the carriers in the electron region in a diode of p germanium is also 2×10^{19} cm^{-3}.

	n, cm^{-3}	μ_n	p, cm^{-3}	μ_n
Diode of n germanium	$2 \cdot 10^{19}$	0.5 kT	$3.5 \cdot 10^{19}$	5 kT
Diode of p germanium	—	—	$3.5 \cdot 10^{19}$	5 kT

As regards the second pair of diodes of n and p germanium with identical value $V_{max} \simeq$ $\simeq$ 70 mv, the density of the tunnel diode at the maximum is different for them; the density of the current at the diode made of p germanium is greater (2×10^2 amp/cm^2) than for the diode of n germanium (8 x 10 amp/cm^2). As shown in [1], the density of the tunnel current depends essentially on the concentration of the majority carriers in the n region, with increasing concentration in the n region the density of the tunnel current increases; the voltage corresponding to the maximum tunnel current remains unchanged here.

In connection with the foregoing, it is necessary to assume that in a diode of p germanium the concentration of the majority carrier in the n region is greater than in a diode of n germanium. Since we know that the diode of n germanium has a majority carrier concentration of 2×10^{19} cm^{-3} in the n region, the diode of p germanium should have in the n region a carrier concentration greater than 2×10^{19} cm^{-3}. From a comparison of the characteristics of the two diodes from p-type germanium, with $V_{max} = 80$ mv ($p = 3.5 \times 10^{19}$ cm^{-3}) and $V_{max} = 70$ mv ($p = 2.5 \times 10^{19}$ cm^{-3}) we can conclude that with decreasing majority carrier concentration in p germanium the concentration of the electrons in the n region increases. It is obvious that when a drop of indium with arsenic is diffused into the p germanium, the arsenic more than offsets the gallium present in the germanium, so that an n region is produced. Then, actually, with the narrow p-n junction produced under identical conditions, the concentration of the majority carrier in the n region should increase with decreasing concentration of gallium in the germanium, and vice-versa with the increase of concentration gallium in the germanium the concentration of the electrons in the n region should decrease. If the assumed compensation takes place, there should exist such a limit of gallium concentration in the initial germanium, for which the infusion of arsenic can no longer compensate the present acceptors, and no n region is produced.

Actually, we did not succeed in our experiment to obtain tunnel diodes at majority carrier concentration in the p-type germanium greater than at 6×10^{19} cm^{-3}.

It must be noted that by taking p-type germanium as the initial material we can apparently obtain tunnel diodes with better parameters compared with tunnel diodes made with n-type germanium. This is explained, in particular, by the fact that the concentration 6×10^{19} cm^{-3} is not the solubility limit for gallium in germanium, and therefore it is easier to obtain by infusion a more perfect p-n junction structure, than by using n germanium with an arsenic concentration close to the solubility limit.

2. MAXIMUM TUNNEL CURRENT

It was shown in [1] that the maximum tunnel current depends very strongly on the concentration of the majority carriers in the n region; at the same time the voltage corresponding to the maximum remains practically constant (for the same degree of doping of the p region). In this case we can assume that the main influence on the tunnel current is exerted by the transparency of the potential barrier, which depends particularly strongly (exponentially) on the thickness of the barrier, and consequently on the concentration of the majority n and p carriers. In order to test this hypothesis, we compared our experimental data with the expression for the barrier transparency. The latter, as follows from [3], in the case of the simplest case of an homogeneous field in the junction, has the form

$$P \sim \exp\left(-\frac{\pi m_r^{*1/2} \Delta E^{3/2}}{2\hbar e E}\right),$$

where ΔE — width of the forbidden band; E — field intensity in the junction; m_r^* — reduced effective mass. Assuming $\Delta E = e\varphi_c$, where φ_c is the contact potential difference and $E = \varphi_c/w$, where w is the thickness of the blocking layer, equal to

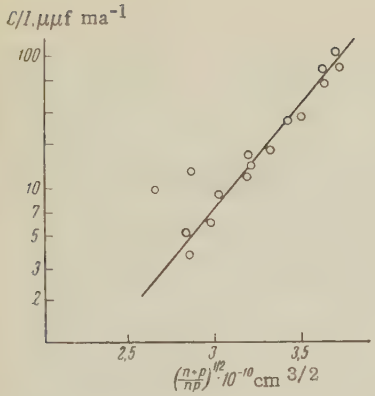


Fig. 3. Dependence of $\frac{C}{I}$ on $(\frac{n+p}{np})^{1/2}$ for diodes made of n germanium with concentration 2.5×10^{19} cm^{-3} in the p region.

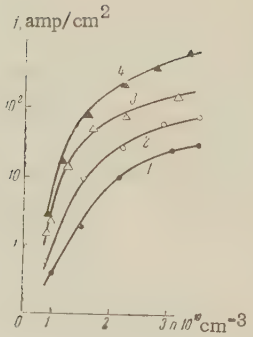


Fig. 4. Dependence of density of tunnel current at the maximum on the concentration of the majority n and p carriers. Concentration in the p region is: 1 — $1.3 \cdot 10^{19}$ cm^{-3}; 2 — $1.8 \cdot 10^{19}$ cm^{-3}; 3 — $2.5 \cdot 10^{19}$ cm^{-3}; 4 — $3.2 \cdot 10^{19}$ cm^{-3}.

$$W \sim \left[\frac{\varepsilon\varphi_K (n+p)}{2\pi e n p}\right]^{1/2},$$

we obtain

$$P \sim \exp\left[-\frac{\pi^{3/2} m_r^{*1/2} \Delta E}{2\hbar\varepsilon^{1/2}} \left(\frac{n+p}{np}\right)^{1/2}\right].$$

To eliminate the unknown junction area we can take the ratio of the junction capacitance C to the tunnel current at the maximum, I. Then

$$C/I \sim \exp A \left(\frac{n+p}{np}\right)^{1/2},$$

where

$$A = \frac{\pi^{3/2} m_r^{*1/2} \Delta E}{2\hbar\varepsilon^{1/2}}.$$

i.e., we should expect a linear dependence of $\ln(C/I)$ as a function of $[(n+p)/np]^{1/2}$. Fig. 3 shows the experimental dependence of $\ln(C/I)$ on $[(n+p)/np]^{1/2}$ for diodes with a like degeneracy in the p region. The line on the figure has a slope 5×10^{10}. The value for $A = 5 \times 10^{10}$ with $\Delta E = 0.7$ ev yields $m_r^* = 0.08$ ma.

Thus, the behavior of the tunnel current at the maximum of the investigated region of concentrations can be approximately described by the expression for the probability of tunnel penetration of the electrons through the potential barrier.

In accordance with the foregoing, the tunnel current at the maximum is quite sensitive to the degree of alloying of the n and p regions. Using relations (1) and (2), we can construct a family of experimental curves for the dependence of the tunnel current density on the concentration of the majority n and p carriers for diodes made of n-type germanium. Fig. 4 shows such curves.

3. DIFFUSION OF IMPURITIES DURING THE INFUSION PROCESS

During the preparation of narrow p-n junctions of p-germanium, it was observed that the maximum of the tunnel current is strongly dependent on the temperature of infusion of indium with arsenic in the germanium. Fig. 5 shows how the maximum of the tunnel current shifts toward lower voltages with increasing infusion temperature. As follows from the statements made above, such a shift in the maximum corresponds to a decrease in the concentration of the carriers in the p region, in spite

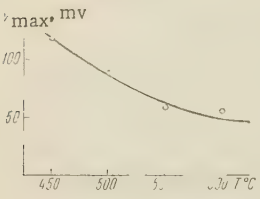


Fig. 5. Voltage corresponding to maximum tunnel current as a function of the infusion temperature ($p = 5 \times 10^{19}$ cm^{-3})

of the fact that for all diodes the initially used germanium had the same concentration p. Such a change in carrier concentration in the p region with changing infusion temperature is possible because of the diffusion of the arsenic during the infusion process. It is known that the rate of diffusion and the solubility of the donor impurities greatly depend on the temperature [4]. Consequently, other conditions being equal (time of infusion and concentration of the arsenic in a drop of indium), the concentration of the donor impurity contained in the p-type germanium increases, thereby more strongly compensating for the gallium present in the germanium, and thus decreases the concentration of the mobile carriers in p-germanium layer situated in the direct vicinity of the p-n junction. The reduction of the concentration in the p region causes the observed change in the voltage corresponding to the maximum tunnel current.

Experiments have shown that as the infusion temperature increases there is also a decrease in the density of the tunnel current, thus indicating a broadening of the blocking layer of the p-n junction with increasing temperature.

In order to reduce the influence of the diffusion, tunnel diodes of p-type germanium were made in vacuum at a minimum possible temperature, 450°C. If the p-n junctions are produced in a hydrogen atmosphere it is possible to use a lower infusion temperature.

CONCLUSION

We investigated experimentally the influence of the degree of alloying of n and p regions of a narrow p-n junction on the tunnel current, and particularly on the magnitude and position of the maximum of the volt-ampere characteristic. For the investigated diodes, which had a large degree of degeneracy in the p-region ($\mu_p > \mu_n$), it was found that the voltage corresponding to the maximum tunnel current is determined by the degree of degeneracy in the hole region: the position of the maximum is proportional to the degree of degeneracy μ_p.

The change in the tunnel current at the maximum with changing majority carrier concentration in the n region is determined essentially by the concentration dependence of the probability of the tunnel penetration of the electrons through the potential barrier.

In conclusion the authors are grateful to S.G. Kalashnikov, V.L. Bonch-Bruevich and N.E. Skvortsov for a discussion of this work.

REFERENCES

1. N.A. Belova, A.N. Kovalev. Radiotekhnika i elektronika. 1961, 6, 1, 160.
2. E. Spenke. Elektronische Halbleiter. Berlin, 1956.
3. F.A. Trumbore. Bell System Techn. J., 1960, 39, 1, 169.
4. L.V. Keldysh. ZhETF. 1957, 33, 4, 994.
5. V. Furukawa, J. Phys. Soc. Japan, 1960, 15, 4, 730.

Institute of Radio Engineering and Electronics,
Academy of Sciences, USSR

Received by editor 14 March 1961

COMPARISON OF STATIC AND DYNAMIC CHARACTERISTICS OF GERMANIUM DIODES UNDER NEGATIVE BIAS

L.S. Lototskii

Results are described on an investigation of germanium diodes under negative bias in the dynamic mode. The design of an experimental parametric microwave amplifier is described. It is indicated that the static characteristic of diodes can be used to design parametric amplifiers.

INTRODUCTION

Semiconductor diodes operate in parametric amplifiers at pumping voltages of considerable magnitude compared with the amplified signal, i.e., in the dynamic mode. At the same time, the parameters of the diodes are determined as a rule at low values of microwave power [1 and 2], i.e., in the static mode.

In the design of parametric amplifiers it is usually assumed that the static and dynamic parameters of the diodes are identical, although this is not obvious. From the data available in the literature one cannot draw any unique conclusions, since the multiplicity of structural features of the parametric amplifiers make it difficult to express their parameters (gain, bandwidth, noise figure) in terms of the static parameters of the diodes. The correctness of this hypothesis cannot be readily checked by means of experiments on complicated parametric amplifiers.

In this connection we undertook the following tasks: first, develop a parametric amplifier with simplest equivalent circuit, the parameters of which can be expressed in terms of the static chracteristics of the diodes; second, by comparing the experimental results with calculation, asertain the possibility of using the conventional procedures to determine the parameters of diodes in it for parametric microwave amplifiers.

1. EQUIVALENT CIRCUIT AND CALCULATION OF PARAMETERS OF EXPERIMENTAL AMPLIFIER

Reference [3] contains the design of a single-loop parametric amplifier with a circulator at the input. The equivalent circuit of such an amplifier is shown in Fig. 1. In this circuit X_L—reactance of the cavity; L—inductance of the diode whisker; C_0—capacitance of the barrier layer of the diode at the operating points; R_l—loss resistance of the diode; C_p—parasitic capacitance due to the cartridge of the diode and to the construction of the cavity; e—signal emf; Z_0—wave impedance of the signal line.

The circuit of Fig. 1 differs from the circuit used in [3] for calculation in the presence of the parasitic capacitance C_p. The influence of this capacitance can be taken into account by replacing the parallel connection of the elements between points a and b of the circuit by an equivalent series connection. This connection gives the circuit shown in Fig. 2, where

$$X = \frac{\omega LD - (D/\omega C_0) = \omega C_p R_1^2}{D^2 + \omega^2 C_p^2 R_1^2}; \quad (1)$$

$$R = \frac{R_n}{D^2 + \omega^2 C_p^2 R_1^2}; \quad (2)$$

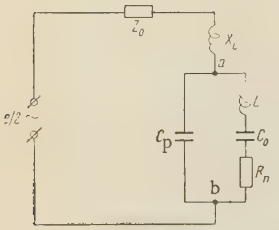


Fig. 1. Equivalent circuit of single-loop parametric amplifier

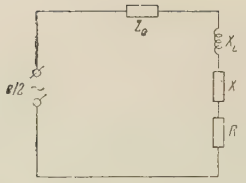


Fig. 2. Equivalent circuit of single-loop parametric amplier with account of the transforming action of the parasitic capacitance C_p.

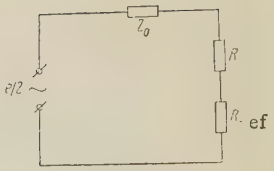


Fig. 3. Equivalent circuit of single-loop parametric amplifier under resonance condition and with the pumping voltage applied to the diode.

and D is a transformation coefficient that takes into account the influence of the parasitic capacitance C_p,

$$D = 1 + (C_p/C_0) - \omega^2 L C_p. \tag{3}$$

When the circuit is tuned to resonance we have

$$X + X_L = 0 \tag{4}$$

and the equivalent circuit of Fig. 2 is transformed to that shown in Fig. 3, where R_{-ef} is the effective negative resistance introduced into the circuit when the pumping voltage is applied to the diode. With account of the transformation we get

$$R_{-ef} = m(1 - a)/2\omega_0 C_0 D^2, \tag{5}$$

where

$$a = \omega^2 L C_p;\ m = \frac{C_{max} - C_{min}}{C_{max} + C_{min}}$$

m—depth of modulation of the capacitance of the barrier layer of the diode.

For the parametric amplifier described by the equivalent circuit of Fig. 3, we can write the expressions for

1) the power gain G:

$$G \simeq \frac{Z_0^2}{\{Z_0 + (R_1/D^2) - [m(1-a)/2\omega_0 C_0 D^2]\}^2}; \tag{6}$$

2) the bandwidth $\Delta f/f_0$:

$$\Delta f/f_0 \simeq \omega_0 C_0 Z_0 D/\sqrt{G}; \tag{7}$$

3) the noise figure F

$$F = 1 + \frac{R_1}{D^2 Z_0}. \tag{8}$$

In reference [3] the calculation is made for an equivalent circuit with lumped circuit parameters L and C. In practice it is impossible to construct for the microwave range a cavity which corresponds completely to the equivalent circuit of [3].

The cavity used in the experimental parametric amplifier of the circuit was tuned to resonance by means of an inductive loop, which modified the frequency characteristic of the amplifier. However, the influence of the inductive loop on the bandwidth of the amplifier can be accounted for. As is well known, the bandwidth is determined by equating the active and

reactive components of the loop impedance

$$X\left(\omega_0 + \frac{\Delta\omega}{2}\right) = r \quad \text{for} \quad X(\omega_0) = 0.$$

In our case

$$X(\omega_0) \simeq (1/D)[\omega_0 L - (1/\omega_0 C_0) - \omega_0 C_p R_1^2/D] + Z_0 \operatorname{tg} \frac{\omega_0 l}{c}, \tag{9}$$

where l—length of the inductive loop, and c—velocity of light,

$$r \simeq Z_0 + (R_1/D^2) - [m(1-a)/2\omega_0 C_0 D^2]. \tag{10}$$

Formulas (6)—(10) are written for the condition $D^2 \gg \omega^2 L^2 C_p^2$. In the microwave band this relation is certainly satisfied, and $\omega^2 L^2 C_p^2$ can therefore be neglected.

Subjecting ω_0 to an increment $\Delta\omega/2$ and expanding in powers of $\Delta\omega/2$, we obtain

$$\frac{\Delta\omega}{2\omega_0}\left\{(1/D)\left[\omega_0 L_0 + (1/\omega_0 C_0) - \omega_0 C_p R_1^2/D\right] + \left(Z_0\omega_0 l/c \cos^2\frac{\omega_0 l}{c}\right)\right\} =$$
$$= Z_0 + (R_{II}/D^2) - [m(1-a)/2\omega_0 C_0 D^2].$$

The expression for the relative bandwidth of the amplifier, with account of the inductive loop, is therefore

$$\Delta\omega/\omega = \frac{2\{Z_0 + (R_1/D^2) - [m(1-a)/2\omega_0 C_0 D^2]\}}{(1/D)[\omega_0 L_0 + (1/\omega_0 C_0) - \omega_0 C_p R_1^2/D] + \left(Z_0\omega_0 l/c \cos^2\frac{\omega_0 l}{c}\right)}. \tag{11}$$

Equation (11) can be simplified. Recognizing that

$$\sqrt{G} = Z_0/\{Z_0 + (R_{II}/D^2) - [m(1-a)/2\omega_0 C_0 D^2]\}$$

and adding and subtracting $1/\omega_0 C_0$ and $Z_0 \tan(\omega_0 l/c)$ to the denominator of (11), we get

$$\Delta\omega/\omega = \frac{2Z_0}{\left\{(2/D\omega_0 C_0) + Z_0\left[\left(\frac{\omega_0 l}{c}\Big/\cos^2\frac{\omega_0 l}{c}\right) - \operatorname{tg}\omega_0 l/c\right]\right\}\sqrt{G}}. \tag{12}$$

Formula (12) differs from expression (7) for the bandwidth in that it contains an additional term in the denominator, so that bandwidth is decreased. In order to use the formulas given above it is necessary to know the parasitic capacitance C_p. This capacitance depends on the construction of the resonator in which the diode is placed, and may not coincide with the capacitance of the diode cartridge, measured in accordance with the procedure given in [1] and [2].

If the coupling between the resonator and the line delivering this signal is independent of the frequency, and if its magnitude is such that the wave impedance Z_0 of the line is completely included in the active losses of the resonant circuit, then the parasitic capacitance C_p can be measured on the basis of the following considerations.

The VSWR in the signal line, for an amplifier tuned to resonance, is

$$\rho = Z_0/R = Z_0 D^2/R_1. \tag{13}$$

If we tune the circuit to resonance in the absence of the pumping frequency at two values of negative voltage applied to the diode, i.e., for two known values of the barrier-layer capacitance C_0 of the diode, then we obtain two values of ρ. Then, using (3) and (13), we obtain two equations for C_p and L:

$$1 + (C_p/C_{01}) - \omega_0^2 LC_p = \sqrt{\rho_1 R_1/Z_0},$$
$$1 + (C_p/C_{02}) - \omega_0^2 LC_p = \sqrt{\rho_2 R_1/Z_0},$$

the solution of which yields

$$C_p = \frac{\sqrt{\rho_2 R_1/Z_0} - \sqrt{\rho_1 R_1/Z_0}}{\frac{1}{C_{02}} - \frac{1}{C_{01}}},$$

where ρ_1, ρ_2, and C_{01}, C_{02} are the values of the VSWR and the barrier-layer capacitance of the layer at bias voltages U_1 and U_2 respectively.

Thus, knowing C_p, we can calculate the gain, the bandwidth, and the noise figure of a parametric amplifier using the static parameters of the diode C_0, R_l, and m.

2. EXPERIMENTAL RESULTS

We designed a parametric amplifier in accordance with the equivalent circuit given above. In the design of this amplifier the characteristic impedance of the line Z_0 was so chosen as to completely compensate for the tuned-circuit insertion losses in the presence of maximum utilization of the diode, i.e.,

$$Z_0 + (R_1/D^2) = m_{\max}(1 - a)/2\omega_0 C_0 D^2. \qquad (14)$$

The plots based on (14) and shown in Fig. 4 allow us to determine the required value of Z_0 for different values of the static diode parameters C_0, R_l, and m. The construction of the amplifier enables us to change Z_0 within the required limits. The bandwidth and the noise figure of the experimental amplifier were measured.

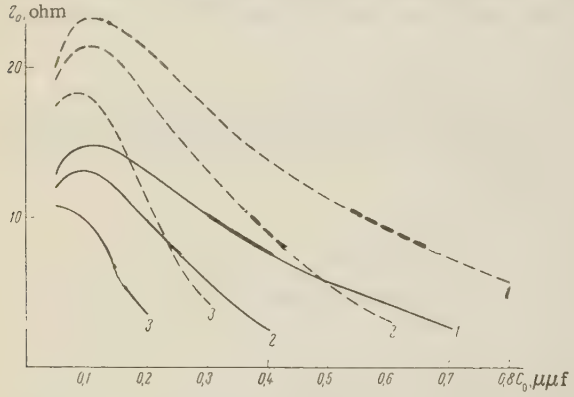


Fig. 4. Z_0 as a function of barrier-layer capacitance C_0 of diode at operating point for different values of m (solid curve, m = 0.4; dashed curve, m = 0.6) and R_e: 1 — R_e = 5 ohms; 2 — R_e = = 10 ohms; 3 — R_e = 20 ohms).

The bandwidth of the amplifier was calculated from the measured gain G and from the known values of Z_0, C_0, ω_0, l, and D. The comparison of the measured gain with the calculated one is meaningless, since the denominator of (6) contains a quantity determined with low degree of accuracy. It must also be noted that the value of a true coefficient of the diode-capacitance modulation is not known exactly for the amplification mode, and this may result in additional errors in the calculation of the gain.

The result of the measurements are listed in the table, which shows the agreement between the measured and calculated bandwidths is satisfactory.

Diode number	C_0, μμf	Measured values		Calculated values
		G, db	Δf, Mcs	Δf, Mcs
444	0,15	27	7	10
468	0,08	27	5	4
455	0,08	29	5	3
360	0,13	22	17	14
470	0,15	25	15	14
378	0,12	25	11	9
357	0,17	27	10	11
649	0,11	27	9	8
641	0,14	26	14	13
643	0,18	27	7	12

This enables us to conclude that the static parameters of the diodes can be used for the calculation of the basic characteristics of microwave parametric amplifiers.

In conclusion, the author considers it his duty to thank K. A. Merkur'ev for help and guidance in the development of the experimental model of the amplifier, and also to Yu. F. Sokolov and N. E. Skvortsova for a discussion and critical remarks concerning this work.

REFERENCES

1. F.S. Rusin, N.E. Skvortsova, Yu.F. Sokolov. Collection: Semiconductor devices and their applications. Izd. Sovetskoye radio, 1958, 3.
2. Yu. F. Sokolov. Radiotekhnika i elektronika. 1961, 6, 3, 399.
3. V.S. Etkin et al. Radioelektronnaya promyshlennost', 1959, 17.

Received by editor 28 March 1961

BRIEF COMMUNICATIONS

TRANSIENT NOISE IN TWO-PATH RADIOWAVE PROPAGATION

A.V. Prosin

It is known that UHF waves propagate in the troposphere along multiple paths. As a result, the received signal is a superposition of two or more oscillations with different amplitudes and phases. In radio relay lines with frequency modulation and frequency sharing, where both tropospheric scatter and line of sight propagation are used, the multiplicity of paths gives rise to transient noise in the communication channels. In the present report we present a procedure for calculating the transient noise occurring in two-path propagation of radiowaves, for the special case when automatic frequency control of the transmitter is used to select the optimum operating point on the phase characteristic of the two-path channel.

Let the transmitter signal have an instantaneous frequency

$$\omega(t) = \omega_0 + \Delta\omega_m u(t), \tag{1}$$

where ω_0—angular carrier frequency; $\Delta\omega_m$—maximum angular deviation of the frequency; u(t)—multipath message satisfying the condition $-1 \leq u \leq 1$ with probability close to unity. The modulating function u(t) is a stationary random process obeying a normal probability distribution law.

In the case of two-path radiowave propagation the frequency phase characteristics of the propagation medium will then be

$$A(\omega) = \sqrt{1 + k^2 + 2k\cos\omega(t)\tau_d}, \tag{2}$$

$$\varphi(\omega) = \operatorname{arc\,tg}\left[\frac{k\sin\omega(t)\tau_d}{1 + k\cos\omega(t)\tau_d}\right], \tag{3}$$

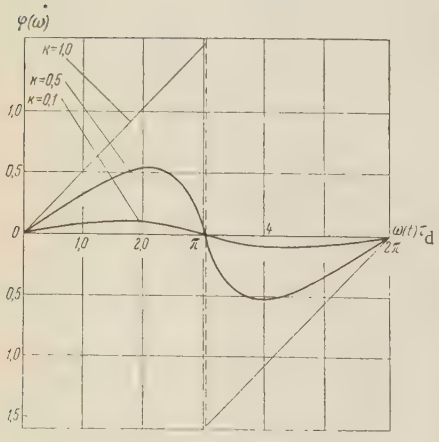


Fig. 1

where $k \leq 1$ is the ratio of the signal amplitudes (smaller to larger); τ_d—relative time delay between two paths.

Fig. 1 shows the phase characteristics of the propagation medium for three values of k. The plots of formula (3) cover the range from 0 to 2π. It follows from (2), (3), and Fig. 1 that the frequency and phase characteristics are periodic functions of the argument $\omega(t)\tau_d$.

An analysis of Eq. (3) and Fig. 1 shows that the cross talk will be minimum when the operating point on the phase characteristic of the propagation medium is a multiple of $q\pi$. At such a choice of operating point the phase characteristic for k = 1 is linear within the limits $\pm\pi$. In this case there will be no transient noise in the communication channels if an ideal frequency detector is used in the receiver.

In practice k and τ_d are random functions of the time, the rate of change of which is quite small

compared with the rate of change of the modulating message. The variations of the ratio of the signal amplitudes will lead to a change in the form of the phase characteristic. The fluctuations of the relative time delay cause the operating point to wander over the phase characteristic of the propagation medium.

It follows from (1) and (3) that the position of the working point on the phase characteristic is determined by the product $\omega_0\tau_d$. Inasmuch as τ_d varies rather slowly compared with u(t), we apparently can, by automatic control of the transmitter carrier frequency, choose and maintain the required operating point on the phase characteristic. A description of practical means for realizing this special automatic control is beyond the scope of the present report. In particular, it can apparently be realized by measuring the amplitudes of the second and higher harmonics, arising when a special test tone is transmitted over the communication lines.

Let us proceed to calculate the transient noise under the condition that this automatic control has chosen on the phase characteristic of the propagation medium the optimum operating point, corresponding to the product $\omega_0\tau_3 = n \cdot \pi$, where $n = 0, 1, 2, 3, \ldots$ $\Delta\omega\tau$, where $\Delta\omega = \Delta\omega_m u(t)$:

$$\varphi(\Delta\omega) = \varphi_0 + \varphi_1\Delta\omega + \varphi_2\Delta\omega^2 + \varphi_3\Delta\omega^3 + \ldots + \varphi_n\,\Delta\omega^n + \ldots . \tag{4}$$

Since the function (3) is odd, the constant term φ_0 in (4) as well as the even terms of the expansion, such as φ_2, φ_4, etc., will vanish. The second term in (4) is the initial undistorted multipath message. Then the expression for the phase distortion will be

$$\Delta\varphi(\Delta\omega) = \varphi_3\Delta\omega^3 + \varphi_5\Delta\omega^5 + \varphi_7\Delta\omega^7 + \ldots . \tag{5}$$

In formulas (4) and (5)

$$\varphi_1 = \frac{\tau_d}{1!}\,\frac{k}{1+k}, \tag{6}$$

$$\varphi_3 = -\frac{\tau_d^3}{3!}\,\frac{k(1-k)}{(1+k)^3} = -\frac{\tau_d^3}{3!}\,y_1, \tag{7}$$

$$\varphi_5 = \frac{\tau_d^5}{5!}\,\frac{k(1-k)}{(1+k)^6}\,(k^6 - 6k^5 - 33k^4 - 52k^3 - 33k^2 - 6k + 1) = \frac{\tau_d^5}{5!}\,y_2, \tag{8}$$

$$\varphi_7 = \frac{\tau_d^7}{7!}\,\frac{k(1-k)}{(1+k)^{12}}\,(1 - 571k^2 - 664k^3 + 498k^4 + 1316k^5 + 850k^6 - 44k^7 - 211k^8 - 6k^9) = \frac{\tau_d^7}{7!}\,y_3. \tag{9}$$

Fig. 2 shows the functions y_1, y_2, and y_3 vs. k. It follows therefore that when k = 1 the functions y_1, y_2, and y_3 vanish. Consequently when k = 1 the expression (5) for the phase distortion also vanishes.

It is shown in [1] that in a telephone channel at a point with zero relative level the psophometric power of the cross talk resulting from nonlinearity of the phase characteristic of a two-port is

$$P_n = 10^9\,\frac{\Delta F_c K_p^2}{\Delta F}\cdot\Omega_c^2\cdot 6\varphi_3^2\Delta\omega_n^4 e^{6b}\,\mathbf{v}_{y_3}(\beta,\sigma) + \ldots . \mu\mu\omega \tag{10}$$

where the phase characteristic of the two-port is approximated by the polynomial (4). Consequently, the calculation of the crosstalk in double-path propagation of radiowaves reduces to a calculation of the transient noise occurring when a multipath frequency-modulated signal is transmitted through a two-port with nonlinear phase characteristic.

Substituting (7)—(9) in (10) we obtain a calculation for the transient noise in two-path radiowave propagation

$$P_{nm} = \frac{10^9}{6}\,\frac{k^2(1-k)^2}{(1+k)^6}\,\frac{\Delta F_n K_n^2}{\Delta F}\,\Omega_c^2\tau_d^6\Delta\omega_n^4 e^{6b}\,\mathbf{v}_{y_3}(\beta,\sigma) + \ldots , \tag{11}$$

where ΔF_c is the bandwidth of the telephone channel; (K_p—psophometric coefficient $K_p = 0.75$

when $\Delta F_c = 3.1$ kcs); $\Delta F = F_2 - F_1$ — bandwidth of all the channels in the linear spectrum of the multipath message; F_2 and F_1 — upper and lower limits of the linear spectrum; $\Omega_c = 2\pi F_c$; F_c—mean frequency of the channel in the linear spectrum of the multi-path communication; $\Delta\omega_c = 2\pi\Delta f_c$; Δf_c—effective frequency deviation per channel; b_{av}—difference in nepers between the average power level of all the channels and the measuring level of one channel (in accordance with the ITCC (Intl. Telephony Consultative Committee) recommendantions [2] $b_{av} = 1.72 + 0.5 \ln N$, where N—number of telephone channels; $y_3(\beta, \sigma)$—a coefficient determined from the curve of Fig. 6.2 of [2].

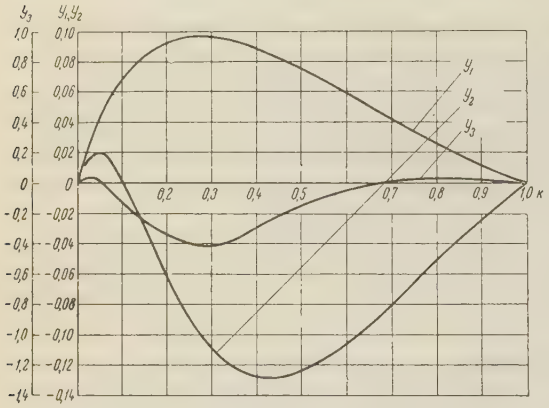


Fig. 2

In real communication systems on UHF the time delay usually does not exceed a fraction of a microsecond. At such values of τ_d it is relatively easy to fulfill the inequality

$$m_{me}\Omega_2\tau_d \leqslant 1, \qquad (12)$$

where m_{me}—effective index of modulation. When (12) is satisfied, it is necessary to take into account in (11) only the noise due to third-order nonlinearity (first term of (11)). Higher-order nonlinearity will be considerably less and can therefore be disregarded. Henceforth in the analysis of transient noise we shall assume that inequality (12) is fulfilled.

The table lists, in accordance with the recommendations of the ITCC [2], the parameters of multichannel radio-relay communication lines with frequency modulation and frequency channel sharing. The last column yields the values of τ_d, calculated on the basis of (12).

Maximum number of working channels	End-point frequency of line spectrum, kc	Effective frequency deviation per channel, kc	Effective index of modulation	Value of β	τ_d, microseconds
60	60-300	200	0.93	5	0.57
120	60-552	200	0.72	9.2	0.401
240	60-1052	200	0.53	17.5	0.286
480	60-2044	200	0.39	34	0.2
600	60-2540	200	0.345	42.3	0.182
900	60-4028	200	0.29	67	0.137

Analysis of (11) shows that the crosstalk has a maximum in the channel with frequency $\Omega_c = \Omega_2$. When k = 0 and k = 1 there will be no crosstalk. The maximum distortions occur when $k_m = 2 - \sqrt{3}$. The maximum transient noise at $k = k_m$, according to (11) and (12), will be

$$P_{nm} = \frac{10^9}{6}\frac{(3\sqrt{3}-5)^2}{(3-\sqrt{3})^6}\frac{\Delta F_k K_n^2}{\Delta F}\Omega_2^2\tau_d^6\Delta\omega_k^4 e^{6b}a v_{y3}(\beta, \sigma). \qquad (13)$$

In (13) the constant coefficient is 1.4×10^6.

In determining the spectrum of the crosstalk [and consequently also in (11) and (13)] the values of k and τ_d were assumed constant, since their rate change is quite small compared with the rate of change of the multipath message. We shall henceforth assume that k and τ_d are independent stationary random processes. Then the power P_{nm} will be, like k and τ_d, a random stationary process, for which the time average is equal to the statistical mean

$$P_{na} = \overline{P_{nm}(k,\ \tau_d)} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P_{nm}(k, \tau_d)\ W\ (k,\ \tau_d)\ dk\, d\tau_d. \tag{14}$$

We assume further that: 1) the amplitudes of two signals arriving at the reception point are independent and satisfy a Rayleigh probability distribution law, having equal mean values,

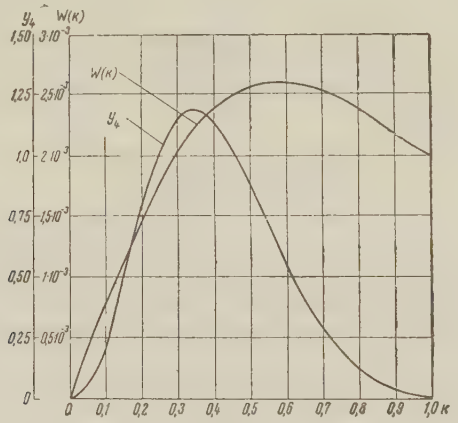


Fig. 3

2) the relative delay is uniformly distributed in the interval τ_{dm}, where τ_{dm} is the maximum time delay. Then the distribution functions of k and τ_d will be equal to [Eq. (15) is derived in the Appendix]:

$$W\ (k) = \frac{4k}{(1+k^2)^2},\quad 0 \leqslant k \leqslant 1,\quad W\ (k) = 0,\quad 0 > k > 1, \tag{15}$$

$$W\left(\frac{\tau_d}{\tau_{dm}}\right) = 1,\quad 0 \leqslant \frac{\tau_d}{\tau_{dm}} \leqslant 1,\quad W\left(\frac{\tau_d}{\tau_{dm}}\right) = 0,\ 0 > \frac{\tau_3}{\tau_{3m}} > 1. \tag{16}$$

Substituting (11) and (15) and (16) in (14) and taking (12) into account we obtain

$$P_{na} = \frac{2}{3}\,10^9\,\frac{\Delta F_c K_p^2}{\Delta F}\,\Omega_2^2\tau_{dm}^6\Delta\omega_H^4 e^{6b}av y_3\,(\beta,\ \sigma)\int_0^1\int_0^1 \frac{k^3\,(1-k)^2}{(1+k)^6\,(1+k^2)^2}\left(\frac{\tau_d}{\tau_{dm}}\right)^6 \times$$
$$\times\, dk d\left(\frac{\tau_d}{\tau_{dm}}\right) = 10^5\,\frac{\Delta F_c K_p^2}{\Delta F}\,\Omega_2^2\Delta\omega_c^4\tau_{dm}^6 e^{6b}av y_3\,(\beta,\ \sigma). \tag{17}$$

Fig. 3 shows a plot of the function $y_4 = k^3\,(1-k)^2(1+k)^{-6}(1+k^2)^{-2}$. From an analysis of (13) and (17) it follows that when averaging over k and τ_d, the power of the transient noise decreases by 14 times compared with the maximum (13). It is interesting to note that in the case of uniform distribution of k within a range from zero to unity, the power P_{ns} is also determined in practice by (17).

In conclusion we give an example of the calculation of the transient noise. A multipath system has the following parameters: N = 240, ΔF_c = 3.1 kc, K_p = 0.75, ΔF = 1052—60 kc, F_2 = 1052 kc, Δf_c = 200 kc, b_{av} = 1, τ_{dm} = 0.2 microsecond, $y_3(\beta, \sigma)$ = 0.45. Substituting all the necessary values in (17) we obtain P_{ns} = 200 micromicrowatt.

APPENDIX

Derivation of Formula for the Distribution Function of the Quotient of Two Independent Random Quantities Obeying a Rayleigh Law With Parameters σ_1 and σ_2

The distribution function of the quotient has the form

$$W\left(\frac{x_2}{x_1}=k_1\right)=\frac{k_1}{\sigma_1^2\sigma_2^2}\int_0^\infty x_1^3 e^{-\frac{x_1^2}{2\sigma_1^2\sigma_2^2}\left(\sigma_2^2+\sigma_1^2k_1^2\right)}dx_1. \tag{I}$$

Integrating, we obtain

$$W\left(\frac{x_2}{x_1}=k_1\right)=2\frac{\sigma_1^2\sigma_2^2k_1}{\sigma_2^2+\sigma_1^2k_1^2}. \tag{II}$$

Analogously we have for $x_1/x_2 = k_2$

$$W\left(\frac{x_1}{x_2}=k_2\right)=2\frac{\sigma_1^2\sigma_2^2k_2}{\sigma_1^2+\sigma_2^2k_2^2}. \tag{III}$$

When $\sigma_1 = \sigma_2$ we have

$$W(k_1)\cdot W(k_2)\qquad W(k)=\frac{2k}{(1+k^2)^2}. \tag{IV}$$

In (I) — (IV) the amplitude ratios k, k_1 and k_2 vary from 0 to ∞. Choosing always the ratio of the interfering beams in such a way as to make $k \leq 1$, we obtain

$$W(k)=\frac{4k}{(1+k^2)^2},\quad 0\leqslant k\leqslant 1,\qquad W(k)=0,\quad 0>k>1. \tag{V}$$

REFERENCES

1. S.V. Borodich. Calculation of noise in radio relay channels with frequency sharing and frequency modulation. Elektrosvyaz', 1956, 1, 10.
2. Engineering-Technical Handbook of Electric Communication, VII, Radio Relay Lines, Svyazizdat, 1956.

Received by editor 20 June 1961

CONCERNING ONE CASE OF DETERMINATION OF THE STATISTICAL CHARACTERISTICS OF THE PHASE OF A VACUUM-TUBE SELF-OSCILLATOR IN RESPONSE TO A LOW-AMPLITUDE VOLTAGE

A.N. Shusterovich

Certain interest attaches to the determination of the energy spectrum, the correlation function, and the phase dispersion of the oscillations in a single-circuit self-oscillator of the Thomson type, in response to a small external voltage of constant amplitude, if the deviation of this voltage relative to the self-oscillator frequency is a stationary random process with zero mean value.

It is known that when a small-amplitude voltage is applied to a single-circuit self oscillator of the Thomson type, the phase of the self oscillations can be described [2] by an abbreviated nonlinear differential equation of the form

$$\frac{d\varphi}{dt} + \alpha_0 \sin \varphi = \Delta, \tag{1}$$

where $\alpha_0 = E_0\omega_0/2A_0$ — synchronization band of the self oscillator: $A_0\omega_0$, φ — amplitude, frequency and phase of the self oscillations; $\Delta = \omega_0 - p$ — frequency deviation; E_0 and p — amplitude and frequency of the external voltage; t — time. Introducing the relative time $\tau_1 = t\alpha_0$, expression (1) can be reduced to the form

$$\frac{d\varphi}{d\tau_1} + \sin \varphi = \Delta_1, \tag{2}$$

where $\Delta_1 = \Delta\alpha_0$ — relative frequency deviation.

In view of the difficulty in solving the nonlinear equation (2), it is sometimes replaced by the linear equation

$$\frac{d\varphi}{d\tau_1} + \varphi = \Delta_1. \tag{3}$$

Such a substitution is permissible if $\Delta_1 \ll 1$. A linear system described by equation (3) has a frequency characteristic

$$f(\Omega) = \frac{1}{\sqrt{1 + \Omega^2}}, \tag{4}$$

where Ω — relative frequency (frequency referred to the synchronization band).

Let Δ_1, which is a relative deviation of small magnitude, be a continuous differentiable stationary random process with zero mean value and energy spectrum $f(\Omega)$.

Then the energy spectrum of the phase of the oscillations of the self oscillator $F_\varphi(\Omega)$ can be represented, taking (4) into account, in the form

$$F_\varphi(\Omega) = \frac{F(\Omega)}{1 + \Omega^2} \tag{5}$$

and the correlation function $B\varphi(\tau)$ and the dispersion $B\varphi(0)$ of the phase be represented by [1]

$$B_\varphi(\tau) = \frac{1}{2\pi}\int_0^\infty \frac{F(\Omega)}{1+\Omega^2}\cos\Omega\tau\, d\Omega, \tag{6}$$

$$B_\varphi(0) = \frac{1}{2\pi}\int_0^\infty \frac{F(\Omega)}{1+\Omega^2}\, d\Omega, \tag{7}$$

where τ is the interval of the relative time τ_1.

We note that the integrals (6) and (7) are always convergent, even when the correlation function of the frequency deviation

$$B(\tau) = \frac{1}{2\pi}\int_0^\infty F(\Omega)\cos\Omega\tau\, d\Omega$$

and the dispersion of the frequency

$$B(0) = \frac{1}{2\pi}\int_0^\infty F(\Omega)\, d\Omega$$

are divergent integrals.

Thus, the requirement that the detuning process be continuous is not essential for formulas (5), (6) and (7) to be applicable.

Expressions (6) and (7) are not always easy to integrate. In this connection interest can attach to a determination of the conditions under which we can neglect (3) the derivative, for a specified accuracy of determination of the dispersion of the self-oscillator phase. The corresponding expression greatly simplifies if the derivative is neglected:

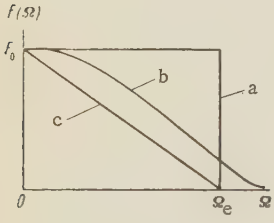
$$B_\varphi(0) \simeq B(0). \tag{8}$$

Obviously, neglect of the derivative in (3) is possible if the frequency deviation Δ_1 varies sufficiently slowly and the phase of the self-oscillator varies slowly.

To take into account the influence of the derivative in (3), we introduce a "slowness coefficient" M:

$$M = \frac{B_\varphi(0)}{B(0)},$$

where $B\varphi(0)$ is the dispersion of the phase of the self oscillator with account of the derivative in (3); B(0) is the dispersion of the frequency deviation Δ_1. In accordance with (8) this dispersion is equal to the dispersion of the phase of the self-oscillator, provided we neglect the derivative in (3). The closer the slowness coefficient to unity, the less the influence of the derivative in (3) on the phase dispersion.



The table shows the values of the slowness coefficient for different forms of frequency-deviation energy spectra (Fig. 1), as a function of the boundary frequency of the spectrum Ω_b. For the spectrum of Fig. 1b, which has the form

$$F(\Omega) = \frac{F_0}{1+a^2\Omega^2},$$

where a = const, we take the end-point to mean the effective spectrum width

$$\Omega_b = \frac{\frac{1}{2\pi}\int_0^\infty F(\Omega)\, d\Omega}{F(0)}. \tag{9}$$

An analysis of the table leads to the following conclusions.

Values of the Slowness Coefficient for Different Energy Spectra of the Frequency Deviation

Form of Spectrum	ω_r					
	0,1	0,2	0,3	0,5	0,8	0,95
Rectangular (Fig. 1a)	1	0,990	0,977	0,935	0,835	0,8
Similar to Fig. 1b	1	0,99	0,985	0,945	0,850	0,860
Triangular (Fig. 1c)	1	1	1	0,98	0,99	0,89

The smallest slowness coefficients are possessed by a rectangular spectrum in which the spectral density is constant within the range $\Omega = 0 - \Omega_b$.

With decreasing spectral density at the high frequencies of the spectrum the corresponding slowness coefficients increase. Thus, for example, for a triangular the slowness coefficients are greater than for the spectrum of the form shown in Fig. 1b.

Accordingly, in real spectra of similar type the slowness coefficients are also greater than in a rectangular spectrum.

In this connection, the slowness coefficients for a rectangular spectrum can serve as a criterion for determining the boundary frequency of the energy spectrum of slow process, i.e., a process in which the derivative can be neglected in Eq. (3) for a specified accuracy in the determination of the dispersion of the self-oscillator phase.

Thus, for example, if the accuracy in the determination of the dispersion permitted is on the order of 5 percent, we obtain from the table that for this case

$$\Omega_b = 0,4 - 0,5.$$

It should be noted that in choosing Ω_b by this method for real spectra, we determine this value with a "safety factor", since actually the permitted boundary frequency is higher.

We can also assume that the processes which are slow in the sense indicated above for Eq. (3), are also slow for Eq. (2), provided the probability of having $\Delta_1 > 1$ is very small. In this case we can determine the dispersion of the self-oscillator phase neglecting the derivative in (2).

REFERENCES

1. B.R. Levin. Theory of random processes and its applications in radio. Soviet Radio Press, 1960.
2. R.V. Khokhlov. On the theory of locking a self-oscillator by an external signal of small amplitude. Proceedings, Academy of Sciences USSR. 1954, 43, 3, 41

Received by editor 4 April 1961

HIGH-FREQUENCY ROTATING FIELD IN CYLINDRICAL CAVITY

L.I. Artemenkov, I.V. Galkin

As shown in reference [1], it is possible to produce effective confinement of a plasma in a constant longitudinal field by using electromagnetic standing waves of circular polarization.

A high-frequency field of this type can be obtained in a cylindrical cavity operating in the H_{111} mode under the excitation conditions indicated below.

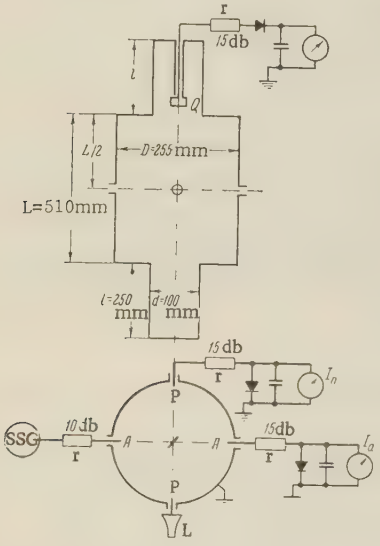


Fig. 1. Diagram of the experiment: A-A — direction coinciding with the orientation of the rod coupled to the generator — active polarization; P-P — direction of passive polarization, turned 90° relative to the active polarization; R — attenuator; L — matched load; Q — indicator loop.

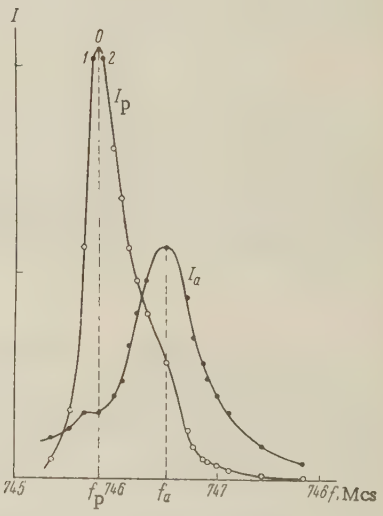


Fig. 2. Resonance curves for two polarizations in the case when the resonant frequencies of the passive (f_p) and active (f_a) polarizations do not coincide ($f_p \neq f_a$).

The experiments were made with a copper cylinder called cavity (Fig. 1) at a supply generator wavelength $\lambda_g \approx 40$ cm. The fields in the cavity were determined by means of loops and two rods. The loop was placed in the beyond-cutoff segment of the round waveguide on the end of the cavity and could be rotated about the axis of the latter. The loop dimensions (~ 10 x 10 mm) and the distance from the plane of the end of the cavity were chosen such that the presence of the loop did not distort the field pattern inside the cavity.

The rods were oriented in the direction of the electric field of the active and passive polarizations. Their depth of insertion could be varied.

The experiments performed have shown that when the resonant curves of both polarizations overlap (Figs. 2 and 3)* and when the the tuning to the frequency f_p is exact, the dependence of the loop detector on the angle of rotation is nearly elliptical (Fig. 5, curve O) regardless of the detuning relative to f_a. In this case the access of the ellipse are oriented along the preferred directions (A—A; P—P) with accuracy of $\pm$ 10°.

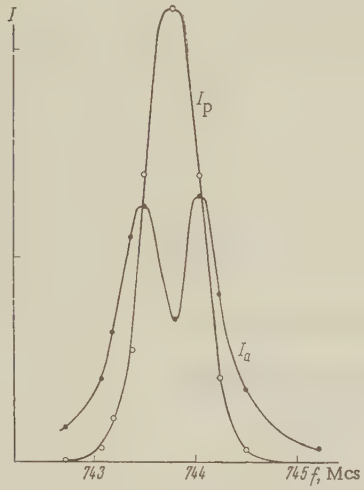


Fig. 3. Resonant curves for two polarizations in the case when $f_p = f_a$.

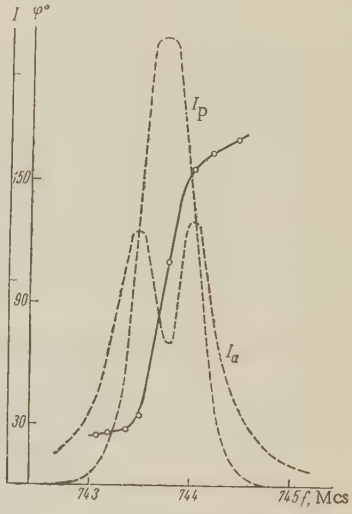


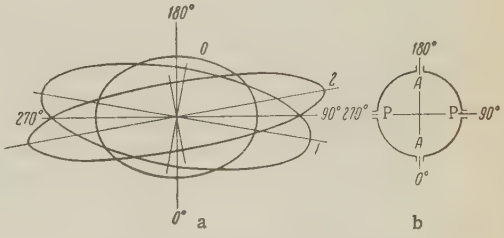
Fig. 4. Example of dependence of the angle of rotation of the axis of the ellipse on the detuning for the case $f_p = f_a$.

The ratio of the axes of the ellipse can be regulated by changing the Q in each of the polarizations and the detuning relative to f_a, so that we can obtain from the readings of the detector coupled to the loop an ellipse that is quite close to a circle.

In the case of negative detuning relative to f_p (point 1 on Fig. 2), the axis of the ellipse shifts clockwise and when the detuning is positive (point 2 on Fig. 2) it shifts counterclockwise. The ratio of the semiaxis is different here for different cases.

Fig. 5

a — dependence of the loop detector current on the angle of rotation for the points 0, 1, 2 of the resonant curve of passive polarization (Fig. 2);
b — position of resonator relative to the polar coordinates.



The change in the angle of rotation of the axes of the ellipse as a function of the detuning, shown in Fig. 4, when the detuning is small and consequently when the field intensity changes little, is due to the change in the phase shift between the oscillations of the active and passive polarizations.

*In Figs. 2, 3, and 4 the ordinates are the relative currents in the rod detectors, and different scales are used for I_p and I_a.

All the foregoing enables us to conclude that when the generator is tuned to the resonant frequency of passive polarization between the field components of the H_{111} mode, corresponding to two preferred mutually-perpendicular directions, a time phase shift of $\pi/2$ is produced.

This result corresponds qualitatively to a $\pi/2$ phase shift between the primary and secondary currents in inductively coupled circuits with lumped parameters when the second circuit is tuned to the generator frequency.

REFERENCES

1. R.Z. Sagdeev. Collection, Plasma physics and problem of controllable thermonuclear reactions, edited by Academician M.A. Leontovich. III, p. 346, AN USSSR, 1958.

Received by editor 3 March 1961

MEASUREMENT OF THE POWER OF AMPLITUDE FLUCTUATIONS OF REFLEX KLYSTRON

S.A. Kornilov, M.V. Lavrov

The simplest method of measuring the amplitude fluctuations of a microwave generator is direct detection [1, 2]. When the investigated oscillations have limited power, the only detector that can be used at present is a semiconductor diode. The use of this diode, however, is difficult because the excess noise increases strongly in the low-frequency region, the spectral density of the noise varying as 1/f.

It is possible to separate the detected generator noise from the internal noise of the detector by using coherent (balance) measurements, the use of which for the investigation of microwave generator noise is described in references [3] and [4]. In particular, Nikonov [4] published the results of measurements of the amplitude fluctuations of reflex klystrons for the 3-cm band in a range from 100 kc to 30 mc. In the present communication we report the preliminary results of measurements of the power of amplitude fluctuations of a reflex klystron in the range from 5 to 200 kc. Since the excess noise of the crystals is quite considerable at these frequencies, the measurements necessitated the construction of a coherent measuring system, capable of registering the power of the detected amplitude fluctuations at a level 20 db below the power of the internal noise of the detectors.*

As in [4], the measurements were made in the 3-cm band.

The block diagram of the measuring setup is shown in Fig. 1. The investigated microwave oscillations were split among two channels by a bridge unit. After detection the fluctuations are amplified and fed to a switching unit, which produces at its output the sum or the difference of the channel voltages.

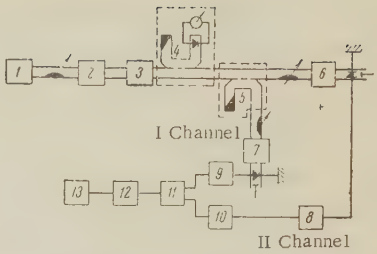
If the investigated generator is not noise modulated, then the in-phase or out-of-phase connection of the channels yields identical readings of the output power meter, since the internal noises of the channels are incoherent. In the case of amplitude modulation of the generator, detection will produce in both channels a coherent component, which causes a

---

*The methodological problems connected with two-channel measurements were solved jointly with S.S. Kratetskii and E.I. Khatskevich.

Fig. 1.

1 — investigated generator; 2 — ferrite modulator; 3, 6, 7 — ferrite gates; 4 — power indicator; 5 — matched slotted bridge; 9, 10 — broadband channel amplifiers (5 — 200 kc); 8 — phase shifter for balancing the channel phases; 11 — switching unit; 12 — tunable filter for the analysis of the fluctuation spectrum; 13 — meter for the fluctuation power (square-law detector with integrating network and pointer meter).



a difference in the readings of the power meter when adding and subtracting the channel voltages. To prevent errors due to parasitic high-frequency coupling between channels, resulting from the noise modulation of the waves reflected from the detector, ferrite gates are connected in the detector arms of the bridge. The total decoupling between detectors, including the decoupling between the bridge arms, was not less than 50 db in this apparatus.

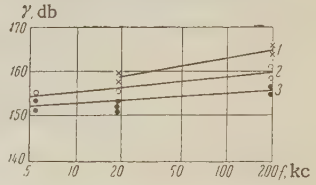
An analysis of the fluctuation spectrum is made after commutation of the signal, so that less stringent requirements had to be imposed on the complete equality of the channel amplifier characteristics; these amplifiers had sufficient bandwidth to transmit the entire investigated fluctuation range.

The filter passband was 300 cps at frequencies from 5 to 20 kc and 2 kc at frequencies from 20 to 200 kc.

The time constant of the integrating circuit of the power meter was chosen on the basis of 100 percent error in the measurement of the noise power, due to natural fluctuations.

Fig. 2. Dependence of the ratio of the signal power to the spectral power density of the amplitude fluctuations on the frequency:

1 — center of the oscillation zone; 2 — half-power point on the low frequency edge of the oscillation zone; 3 — half-power point on the high-frequency edge of the oscillation zone.



The apparatus was balanced (i.e., the amplitudes and phases of the output of the channels in the narrow filter band were suitably set) by external modulation of the microwave oscillations with the aid of a ferrite modulator.

The apparatus was calibrated by comparing the action of low-frequency noise at the input of the amplifiers with the action of sinusoidal voltage of equivalent power, with subsequent recalculation to high frequencies using the conversion losses measured under the experimental conditions.

Fig. 2 shows the results of the measurements of the amplitude fluctuations of a reflex klystron at the center of the oscillation zone and at the half-power points.

In the center of the zone the noise modulation level is lower at the edges, and particularly large noise is observed when the detuning is towards the high-frequency side. With decreasing frequency of analysis, the relative level of the coherent components in the channels decreases, since the spectral density of the amplitude-fluctuation power of the reflex klystron changes much more slowly than 1/f, while the noise spectrum of the detector varies as 1/f. Thus, at the 5 kc point the sensitivity of the apparatus was no longer sufficient to measure the fluctuation power at the center of the oscillation zone.

The results obtained at 200 kc agree with the Nikonov data [4].

Worthy of particular attention of the frequency dependence of the spectral density of the investigated amplitude fluctuations. Our data contradict the data given in the article by Malakhov [5], according to which the low-frequency amplitude fluctuations of a reflex klystron have a 1/f power spectrum.

If this were true, the difficulties in the measurement of the high-frequency amplitude flucutations would not occur: the ratio of the power of the detected noise of the klystron to the power of the internal noise of the detector would remain constant. We are inclined to assume that the frequency dependence we obtained for the amplitude noise of the klystron is overestimated as a result of the inaccuracy in the calibration of the setup, and that actually the

amplitude fluctuations have an almost constant spectrum in the investigated range. Measurements made by Aitchison [6] with a superheterodyne procedure show precisely such a result.

REFERENCES

1. R. Mueller. IRE Trans. 1954, ED-1, 4, 42.
2. W. Gottshalk. IRE Trans. 1954, ED-1, 4, 94.
3. N. Smith. Proc. IEE 1958, B105, Suppl. N. 11, 800.
4. V.N. Nikonov. News of the Universities, Ministry of Higher Education USSR — Radiophysics, 1959, 2, 6, 915.
5. A.N. Malakhov. Ibid, 1960, 3, 6, 1001.
6. C.S. Aitchison. Proc. IEE, 1958, B105, Suppl. N 12, 944.

Received by editor 10 May 1961

CONDITIONS FOR OBTAINING MINIMUM PULSATIONS OF ELECTRON BEAMS FOCUSED BY A HOMOGENOUS MAGNETIC FIELD IN SYSTEMS WITH PARTIALLY SCREENED CATHODE

S.I. Molokovskii

The problem of reducing the pulsations of the electron beam by adjusting the flux of the magnetic field through the surface of the cathode was examined by Mendel [1]. He showed that the pulsations of the flux are minimal if

$$K = 1 - \frac{\omega_p^2}{2\omega_L^2}, \tag{1}$$

where $K = (N_1/N_2)^2$ — screening parameter of the cathode, equal to the square of the ratio of the flux of the magnetic field through the cathode surface to the flux of the magnetic field at the input in the region of the focusing field; $\omega_p = \sqrt{\eta\rho/\epsilon}$ — plasma frequency; $\omega_L = \frac{1}{2}\eta B$ — Larmor frequency.

Relation (1) has been obtained under the assumption that the radial velocities of the electrons at the entrance to the region of the homogeneous field is zero. Since under real conditions one of the many sources of electron-beam pulsations are the non-zero radial velocities of the electrons at the entrance to the focusing zone, it is interesting to determine the condition under which minimum pulsations are obtained, with account of the initial radial velocities of the electrons. This can be determined by finding the minimum of the amplitude of the pulsations as a function K.

Using the results of [1], the amplitude of the pulsations can be represented in the form

$$\delta_m = \sqrt{\frac{\frac{1}{r_0^2}\left(\frac{dr}{dt}\right)_0^2\left[\omega_L^2(1+3K)+\frac{\omega_p^2}{2}\right]+\left[\omega_L^2(1-K)-\frac{\omega_p^2}{2}\right]^2}{\left[\omega_L^2(1+3K)+\frac{\omega_p^2}{2}\right]^2}}, \tag{2}$$

where r_0 is the initial radius of the electron beam; $(dr/dt)_0$ — initial radial velocity of the peripheral electron.

From this we determine, using the conventional method of founding the extremum of a function, the following condition for the minimum of the pulsation amplitude

$$K = 1 - \frac{\omega_p^2}{2\omega_L^2} + \frac{\frac{3}{2}\left(\frac{dr}{dt}\right)_0^2 \omega_L^2 \frac{1}{r_0^2}}{1 + 3\,\frac{\left[1 - K - \frac{\omega_p^2}{2\omega_L^2}\right]}{\left[1 + 3K + \frac{\omega_p^2}{2\omega_L^2}\right]}}. \qquad (3)$$

In the case of zero initial radial velocities this condition, as can be readily seen, goes over into the Mendel condition (1).

Expression (3), together with the initial equation (2), is strictly valid within the framework of the linear theory of electron-beam pulsations [1]. However, numerical calculations on the motion of the peripheral electron, made for the case of large disturbances, using the equations from nonlinear theory [2], yield optimal values of K which are close to the value determined from (3).

REFERENCES

1. J.T. Mendel. Proc. IRE, 1955, 43, 3, 327.
2. C.C. Wang. Proc. IRE, 1950, 38, 2, 135.

Received by editor 20 March 1961

LETTERS TO EDITOR

STORAGE OF NOISE IN AN ACCUMULATOR WITH DELAYED FEEDBACK

In practical realization of optimum filtering of a periodic sequence of pulsed signals in noise, use is made of accumulators with delayed feedback (see [1], Fig. 1a), which comprise a linear filter with transfer function $K(\omega)$, a feedback look with transfer coefficient $m < 1$, a line producing a delay equal to the repetition period T of a pulsed signal, and a summing unit. Under the assumption that all the elements of the accumulator are linear and that all, except the filter, have a uniform amplitude-frequency characteristic within a sufficiently broad band of frequencies, reference [1] has considered the accumulation of white normal noise in such accumulators. In this case the noise power at the output of the accumulator was calculated for the steady state from the frequency characteristics of an accumulator with delayed feedback.

We proposed below a different approach, which enables us to consider the accumulation of noise with non-uniform spectrum and to calculate the noise power at the output of the accumulator with delayed feedback after any number of accumulation cycles.

Let us consider the accumulation of a stationary noise $\nu_0(t)$ with spectra density $S_0(\omega)$ in an accumulator with delayed feedback, $S_0(\omega)$ being such that the noise can be regarded after a time T to be practically independent. We denote by $\nu_m(t)$ the noise $\nu_0(t)$ which has passed n times through the filter of the accumulator. Then the spectra power density of the random process $\nu_n(t)$ will be [2]

$$S_n(\omega) = S_0(\omega) \mid K(\omega) \mid^{2n} \tag{1}$$

From the accumulator diagram it follows that the noise at its output after N repetition periods T will be expression term of the input noise by the formula

$$v(t) = \sum_{n=0}^{N} m^n \nu_{n+1}(t - nT) \tag{2}$$

and its power is

$$\sigma^2(N) = \sum_{n=0}^{N} m^{2n} \sigma_{n+1}^2, \tag{3}$$

where σ_{n+1}^2 — the power of $\nu_{n+1}(t)$ — is calculated from the spectral density (1):

$$\sigma_{n+1}^2 = \frac{1}{2\pi} \int_0^\infty S_{n+1}(\omega)\, d\omega = \frac{1}{2\pi} \int_0^\infty S_0(\omega) \mid K(\omega) \mid^{2n+2} d\omega.$$

Substituting this expression in (3) and interchanging the order of summation and integration, we obtain finally an expression for the noise power at the output of the accumulator with delayed feedback after N repetition periods, if the input is a stationary noise with spectral density of power $S_0(\omega)$:

$$\sigma^2(N) = \frac{1}{2\pi} \int_0^\infty \frac{S_0(\omega) \mid K(\omega) \mid^2}{1 - m^2 \mid K(\omega) \mid^2} \left(1 - m^{2N+2} \mid K(\omega) \mid^{2N+2}\right) d\omega.$$

In particular, in the steady state the noise at the output of the accumulator is obtained from (4) with $N \to \infty$:

$$\sigma^2 = \frac{1}{2\pi} \int_0^\infty \frac{S_0(\omega) \mid K(\omega) \mid^2}{1 - m^2 \mid K(\omega) \mid^2} d\omega.$$

An approximate value of σ^2 was obtained in [1] for white noise under superfluous assumption that it has a normal distribution.

REFERENCES

1. Yu.S. Lezin. Accumulation of noise in devices with delayed feedback. Radiotekhnika i elektronika. 1961, 6, 2, 187.
2. B.R. Levin. Theory of random processes and its application in radio. Soviet Radio Press, 1960.

V.G. Kononov

Received by editor 20 June 1961

RADIO ENGINEERING AND ELECTRONIC PHYSICS

The English Edition of Radio Engineering and Electronic Physics is mailed to subscribers within 18 weeks after the publication of the original Russian issue.

Russian electronic journals published by the
American Institute of Electrical Engineers
Translated by Royer and Roger, Inc.

	Subscription rates			
	Individuals		Libraries	
	$	£	$	£
Radio Engineering and Electronic Physics	28.50	10	57.00	20
Radio Engineering	14.25	5	28.50	10
Telecommunications	14.25	5	28.50	10

Royer and Roger translates and produces
the following Russian scientific journals:

Biophysics
Entomological Review
Geochemistry
Geodesy and Cartography
Izvestiya, Academy of Sciences of the USSR, Geologic Series
Pavlov Journal of Higher Nervous Activity
Problems of Oncology
Radio Engineering
Radio Engineering and Electronic Physics
Refractories
Sechenov Physiological Journal of the USSR
Soil Science
Telecommunications

Comments and inquiries regarding *Radio Engineering and Electronic Physics* and other translation journals should be sent to:

International Division
Royer and Roger, Inc.

1000 Vermont Avenue, N.W.
Washington 5, D.C.

41 East 28th Street
New York 16, New York

RADIO ENGINEERING AND ELECTRONIC PHYSICS

РАДИОТЕХНИКА И ЭЛЕКТРОНИКА

NUMBER 11 1961

CONTENTS

Yu.I. Vorontsov and K.S. Rzhevkin, Tunnel Diodes in Amplifier Circuits (Review) . . . 1585
A.N. Radchenko and E.T. Mironchikov, Sequential Methods of Correcting Single and Multiple Closely-Spaced Errors in Group Codes . . . 1606
S.A. Akhmanov and R.V. Khokhlov, Transformation of Random Signals in Nonlinear Lines . . . 1613
O.N. Litvinenko, Synthesis of Inhomogeneous Lines on the Basis of an Input Impedance Specified in the Form of a Rational Fraction of the Frequency . . . 1624
P.A. Perepelyatnik, Biharmonic Oscillations in a Self-Oscillator with Delay at a Non-Synchronous Frequency Ratio . . . 1631
A.I. Rutkas, Chain Synthesis of a Reactance Multi-Port Network . . . 1637
Ya.S. Shifrin, Correlation Characteristics of the Field of a Linear Antenna . . . 1643
M.D. Khaskind, Diffraction of Waves on a Slot and on a Ribbon Oriented Perpendicular to an Impedance Plane . . . 1656
L.A. Lyubimov, G.I. Veselov, and N.A. Bei, Dielectric Waveguide With Elliptical Cross Section . . . 1668
V.N. Shevchik and N.I. Sinitsyn, Effect of Variation of Electron Velocities on the Operation of a Backward Wave Tube . . . 1677
V.D. Ivanova and V.S. Mikhalevskii, Plasma Control of Frequency of a Traveling Wave Oscillator . . . 1683
L.A. Birger, Bandwidth and Noise Figure of Tunnel-Diode Tuned Amplifier . . . 1688
B.A. Trigubenko, B.M. Tsarev, Thermionic Properties of Hexaborides and other Interstitial Compounds . . . 1694
I.A. Rezgol', Optimum Operating Points on the Volt-Ampere Characteristic of a Thermoelectric Converter . . . 1699
F.M. Yablonskiy, Current Control of the Ignition of a Discharge in a Dekatron . . . 1706
N.A. Belova and A.N. Kovalev, Some Special Features of the Volt-Ampere Characteristics of Narrow Germanium p-n Junctions . . . 1712
L.S. Lototskii, Comparison of Static and Dynamic Characteristics of Germanium Diodes Under Negative Bias . . . 1717

BRIEF COMMUNICATIONS

A.V. Prosin, Transient Noise in Two-Path Radiowave Propagation . . . 1722
A.N. Shusterovich, Concerning One Case of Determination of the Statistical Characteristics of the Phase of a Vacuum-Tube Self-Oscillator in Response to a Low-Amplitude Voltage . . . 1727
L.I. Artemenkov and I.V. Galkin, High-Frequency Rotating Field in Cylindrical Cavity . . . 1730
S.A. Kornilov and M.V. Lavrov, Measurement of the Power of Amplitude Fluctuations of Reflex Klystron . . . 1732
S.I. Molokovskii, Conditions for Obtaining Minimum Pulsations of Electron Beams Focused by a Homogeneous Magnetic Field in Systems with Partially Screened Cathode . . . 1734

LETTERS TO EDITOR

V.G. Kononov, Storage of Noise in an Accumulator with Delayed Feedback . . . 1736